Alternative Splicing Regulation in Programmed Cell Death and Neurological Disorders: A Systems Biology Approach

ALTERNATIVE SPLICING REGULATION IN PROGRAMMED CELL DEATH AND NEUROLOGICAL DISORDERS: A SYSTEMS BIOLOGY APPROACH

A dissertation presented

by

Qingqing Wang

to

The Committee on Higher Degrees in Systems Biology

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Systems Biology

Harvard University
Cambridge, Massachusetts
April 2013

Dissertation Advisor: Professor Pamela A. Silver Qingqing Wang

ALTERNATIVE SPLICING REGULATION IN PROGRAMMED CELL DEATH AND NEUROLOGICAL DISORDERS: A SYSTEMS BIOLOGY APPROACH

Abstract

Alternative splicing (AS) is a major source of biological diversity and a crucial determinant of cell fate and identity. Characterizing the role of AS regulatory networks in physiological and pathological processes remains challenging. The work presented here addresses this challenge using systems biology analyses of AS regulatory networks in programmed cell death and neurological disorders.

The first study describes a genome-wide screen based on splicing-sensitive reporters to identify factors that affect the AS of apoptosis regulators *Bclx* and *Mcl1*. The screen identified over 150 factors that affect apoptosis through modulating the pro- and anti-apoptotic splicing variants of these apoptosis regulators. This screen revealed a new functional connection between apoptosis regulation and cell-cycle control through an AS network. It also unearthed many disease-associated factors as AS effectors.

The second study describes the functions of the Polyglutamine-binding protein 1 (PQBP1)-mediated AS regulatory network in neurological disorders. PQBP1 is a factor linked to intellectual disability and was unexpectedly identified as an AS effector from the screen described above. We found that PQBP1 influences the splicing of many mRNAs and is associated with a wide range of splicing factors. Depletion of PQBP1 in mouse primary cortical neurons caused defects in neurite outgrowth and altered AS of mRNAs enriched for functions in neuron projection regulation. Disease-mutants of PQBP1 lose associations with splicing factors and cannot complement the aberrant AS

patterns and neuron morphology defects in PQBP1 depleted-neurons. This study revealed a novel function of PQBP1 in AS regulation associated with neurite outgrowth and indicated that aberrant AS underlies the pathology of PQBP1-related neurological disorders.

A final study examines the dynamics of the *Drosophila* Sex-lethal AS regulation network using a combination of experimental tools and mathematical modeling. This study demonstrates that the features of *Sxl* AS regulation have great potentials in building synthetic memory circuits in mammalian cells to track cell fate.

Collectively, this work describes the landscape of three diverse AS regulatory networks in various biological processes. The results and methods presented here contribute to our rapidly advancing knowledge of AS regulation in biology and human disease.

TABLE OF CONTENTS

ACKNOWLEDGEMENT

I would like to express my sincere thanks to the many colleagues, advisors, family and friends who supported me through graduate school and witnessed my transformation from a pure mathematics major to a systems biologist.

First and foremost, I thank my advisor Pam Silver, without whom there is no way I could have gone this far. I greatly appreciate her faith in her students, her encouragement of us to think "deep" and "big" in biology, and her support of us to take on ambitious even risky approaches towards scientific questions. Her tutoring of these many years significantly influenced and shaped the way I think about biology and conduct research. I also want to thank her for her caring about her students. Being an international student, from the moment I joined the lab I have been encouraged strongly by her to work hard on my English, and to watch, feel, and enjoy the American culture. I also owe special thanks to her for the tremendous support I received from her during a difficult time early in graduate school when I was not sure how to pursue my research in biology. Finally, I want to thank her for running such a fantastic lab from where I found my American "family."

I have also benefited greatly from the guidance and help of my Dissertation Advisory Committee members Tim Mitchison, Galit Lahav and Michael Greenberg, who kindly shared their expertise, critiques and resources without reservation. I would also like to thank my Dissertation Examination Committee members Tim Mitchison, Jesse Gray, Jack Szostak, and Stirling Churchman who agreed to review my work and attend my defense out of their busy schedules.

I have benefited from the wisdom and kindness from many colleagues and collaborators. In the Silver lab I owe special thanks to my “big brother” Michael Moore, from whom I received much of my basic biological training: an intense and ferocious lesson in “Biological experiments for dummies” when I popped in lab knowing nothing but doing mathematical proofs. I was also lucky to know Natalie Gilks-Farny, Jessica Hurt and Bill Senapedis from whom together with Michael, I learned a tremendous amount about RNA biology and I thank them for their expertise and patience. I also want to thank Karmella Haynes, Jeffrey Way and Daniel Ducat for all the insights and thoughtful advice I received from them in learning and doing research in synthetic biology.

Outside of the Silver lab, I am especially indebted to Michael Greenberg’s lab for the generous sharing of the precious embryonic mouse neurons and their expertise in neuron culture. Their kindness made it possible for me to conduct experiments in neurons that much of my work relied on. I also want to thank Guillaume Adelmant for his expertise in proteomic analysis using mass-spectrometry, and Jennifer Waters and Lauren Piedmont from the Harvard Nikon Imaging Center who made neuron imaging far less painful for me.

I want to thank the Systems Biology administrative office and lab operations staff who provide a solid foundation for the running of the Systems Biology PhD program and all the researches in the department. I especially want to thank Sam Reed who helped me schedule all DAC meetings and with countless other errands. I also owe special thanks to Kathy Buhl, the Master Lab Coordinator of the Silver lab who made

sure everything worked and tolerated my constant bothering of “URGENT please ASAP” orders.

I want to thank all the wonderful friends who populated my journey through graduate school and made the experience not only enjoyable but also a transformative one. I am especially grateful for the big, weird and probably (but properly) dysfunctional fun family of the “Silverinos,” members of whom are too numerous to name. Time passes and face changes, but the essence of this family stays. The characteristics and personalities from three generations of Silverinos during my time here embraced me deeply and shaped my personality greatly.

Last but not the least, I would like to thank my family. I thank my parents for their unconditional love and constant faith in me. Finally, I want to thank my husband Jeffrey Paulsen, whose love and support I have always depended on.

ATTRIBUTIONS

For Chapter 3, Chapter 4 and Chapter 5, I performed all the work presented in this dissertation with exceptions detailed here. Guillaume Adelmant analyzed the proteins associated with PQBP1 and its two disease mutants with mass spectrometry in Chapter 3, and contributed to the written explanation of the method in the supplemental data of Chapter 3.

For Chapter 2, Michael Moore conducted the genome-wide screen based on splicing sensitive reporters and performed subsequent experimental analysis. Michael Moore also prepared the manuscript. I made all the ASF/SF2 and ASF/SF2 mutants constructs used in this study and performed Western analyses examining ASF/SF2 turnover. Caleb Kennedy developed the Support Vector Machine model for analyzing the genome-wide screen data and contributed to the written and graphical explanation of the method.

CHAPTER 1: INTRODUCTION

MESSENGER RNAS: A DYNAMIC READOUT OF GENETIC INFORMATION

The central dogma of molecular biology states that the genetic information of life is stored in DNA and transferred to messenger RNA (mRNA) for expression as proteins (Crick 1970). mRNA however, is more than a passive intermediate in the expression of genetic information. Instead, multiple mechanisms on mRNA processing enable mRNA to selectively extract and re-organize information from DNA. mRNA provides a dynamic readout of the genetic information stored in DNA in a tissue- and cell type-specific way.

It is now appreciated that the number of protein-coding genes in a genome is not proportional to the degree of complexity of that organism (Lodish et al. 2008). For example, the human genome encodes around 25,000 protein-coding genes, only four times as many as the yeast *Saccharomyces* and comparable to that of the plant *Arabidopsis thaliana*, the nematode *C. elegans* and the fruit fly *Drosophila*. Furthermore, human and chimpanzee genomes show a 98% sequence identity, even though humans exhibit far more complicated cognitive behaviors such as the use of language (Lander et al. 2001; Olson and Varki 2004). These examples illustrate that the phenotypic complexity of an organism is not solely determined by the number of protein-coding genes it possesses. Rather, it has been proposed that the diversity and sophistication of the eukaryotic proteome is augmented by mRNAs via regulatory mechanisms at the transcriptional and post-transcriptional levels.

Regulatory mechanisms at the post-transcriptional level, in particular, contribute to mRNA and proteome diversity in eukaryotes. One of the most prominent mechanisms is precursor mRNA (pre-mRNA) splicing. In most eukaryotes genes are ‘split’: the protein-coding regions (exons) of a gene are interrupted by ‘silent’, non-coding regions

(introns). The maturation of an mRNA transcript requires the excision of introns and the concomitant joining of exons (Gilbert 1978; Matlin et al. 2005). In many cases of multi-exon mRNA transcript maturation there is more than one way to remove introns and join exons, enabling a gene to generate multiple mRNA and protein variants (Wang et al. 2008; Pan et al. 2008) in a process termed alternative splicing. In addition to pre-mRNA splicing, mRNA transcripts are subject to many other regulatory processes, such as control of mRNA stabilization, transport and RNA editing. These mRNA processing steps greatly increase the genome's capacity to generate different mRNA profiles and expand the number of functionally distinct proteins that can be encoded by the genome (Nilsen and Graveley 2010).

Multicellular organisms are typically composed of hundreds of distinct cell types that are specialized in a variety of functions. These cells share largely identical genetic information, but are capable of generating diversified gene expression patterns in response to tissue- and development stage-specific needs. The mRNA profiles are thus of special importance as a dynamic reflection of how genes are expressed and regulated. Moreover, aberrant mRNA profiles are associated with human diseases and could provide valuable information on pathology and potential treatment strategies.

MECHANISM OF PRE-MRNA SPLICING

The exon-intron structure of eukaryotic genes greatly facilitates the generation of diverse mRNA profiles and pre-mRNA splicing is a crucial process in the eukaryotes. Pre-mRNA splicing is carried out by the spliceosome, a large and highly dynamic ribonucleoprotein (RNP) machine. The spliceosome is assembled from five basic small

nuclear RNP (snRNP) elements (U1, U2, U4, U5 and U6) and is dynamically associated with up to 300 additional proteins (Rappsilber et al. 2002; Wahl et al. 2009).

The basic splice sites in eukaryotic pre-mRNAs are generally defined by short consensus sequences (Figure 1-1A) and the spliceosome is assembled onto the pre-mRNA in a step-wise manner on those sites. During initiation of the spliceosome assembly (spliceosomal E complex), the 5' splice site (5'SS) is recognized by the U1 snRNP, the branch point sequence (BPS) by SF1 (splicing factor 1) and 3' splice site (3' SS) by U2AF (U2 auxiliary factor) (Figure 1-1B). Following the formation of the E complex, U2 snRNP docks to the BPS and replaces SF1, forming the pre-spliceosome—complex A. Subsequently, U4/U5/U6 pre-assembled tri-snRNPs are recruited to the A complex, forming the pre-catalytic spliceosome—complex B. The spliceosome then undergoes major conformational changes in which U1 and U4 snRNPs are released, giving rise to the activated spliceosome—the B* complex. The spliceosomal B* complex performs the first catalytic step of pre-mRNA splicing: the 2'-hydroxyl group of the adenine base at the BPS attacks the phosphodiester bond at the 5'SS, generating a free 5' exon and an intron lariat-3' exon. This intermediate together with the remaining snRNPs forms the C complex. A second round of spliceosome conformational change takes place here and the second catalytic step of pre-mRNA splicing is performed: the phosphodiester bond at the 3'SS is attacked by the 3'-hydroxyl of the 5' exon, giving rise to exon ligation and excision of the lariat intron. After this the spliceosome dissociates and disassembles, releasing the spliced mRNA and snRNPs for recycling for the next splicing reaction (Figure 1-1B; Wahl et al. 2009).

Figure 1-1: Consensus sequences of eukaryotic splice sites and step-wise spliceosome assembly on splice sites.

(A) The consensus sequences of splice sites in metazoans and yeasts. Schematics of a pre-mRNA are shown, with green rectangles representing exons and black lines between exons representing introns. 5' splice site is represented by the consensus sequence "GUAAGU" in yeast and "GU" in metazoans; branch point sequence by the sequence "UACUAAC" in yeast and "A" in metazoans; 3' splice site by "AG" in both metazoans and yeasts. The consensus sequences of the 5' splice site, branch point sequence and 3' splice site in yeast are highly conserved but in metazoans are degenerate.

(B) Step-wise assembly of the spliceosome on splice sites in metazoans. Schematics of a pre-mRNA are shown, with green rectangles representing exons and black lines between exons representing introns. "GU" marks the 5' splice site (5' SS), "A" the branch point sequence (BPS) and "AG" the 3' slice site (3' SS). See text.

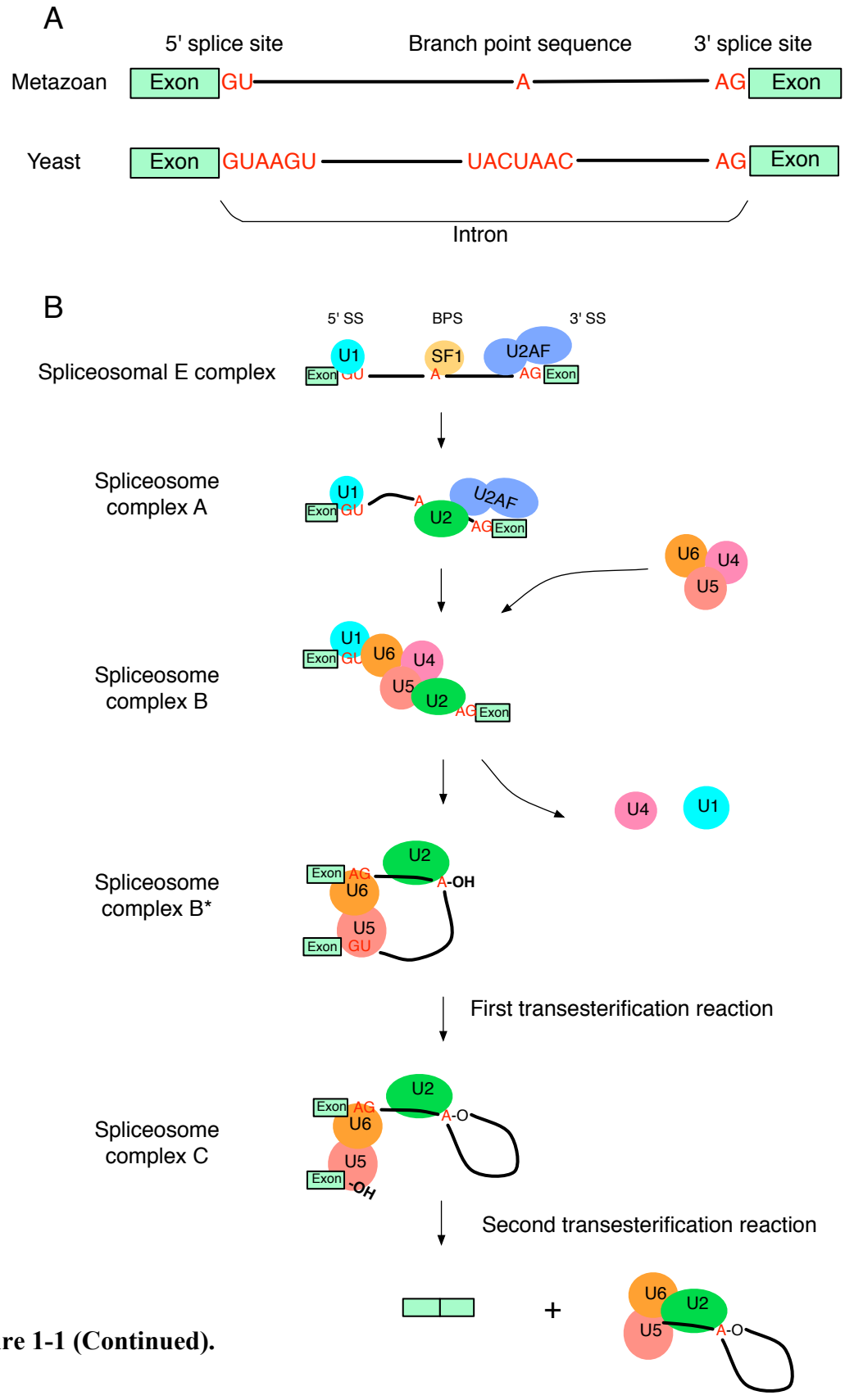

Figure 1-1 (Continued).

ALTERNATIVE SPLICING IN BIOLOGICAL DIVERSITY AND REGULATORY NETWORKS

Alternative splicing (AS) depicts the phenomenon in higher eukaryotes that introns can be removed and exons are pieced together in different patterns to produce functionally distinct mRNA and protein variants from one gene. AS is a major source of biological diversity in eukaryotic transcriptomes and proteomes. The contribution of AS to eukaryotic complexity is likely even greater than previously recognized, as high-throughput studies showed that >90% of mammalian genes undergo AS and generate polymorphic mRNA transcripts (Pan et al. 2008; Wang et al 2008). The number of mRNA transcripts that a multi-exon gene is able to encode varies from several to thousands, a staggering level of diversity. As an example, the *Drosophila* gene Down syndrome cell adhesion molecule (Dscam) is able to make approximately 38,000 mRNA isoforms, far surpassing the number of total protein-coding genes (~14,500) in the organism (Schmucker et al. 2000).

There are seven basic patterns that pre-mRNAs can follow during AS (Figure 1-2; Black 2003). An exon can be alternatively included or excluded in the transcript, and is defined as a cassette exon (Figure 1-2A). Multiple cassette exons can be mutually included or excluded to generate different isoforms (Figure 1-2B). Part of an exon can be alternatively spliced, introducing alternative 5' or 3' splice sites within that exon (Figure 1-2C,D). Furthermore, the use of alternative promoters or polyadenylation (poly-A) sites can also contribute to AS by extending, shortening or changing the first or last exon of the mRNA transcript (Figure 1-2E,F). Finally, an intron that is supposed to be removed can be retained in the transcript (Figure 1-2G).

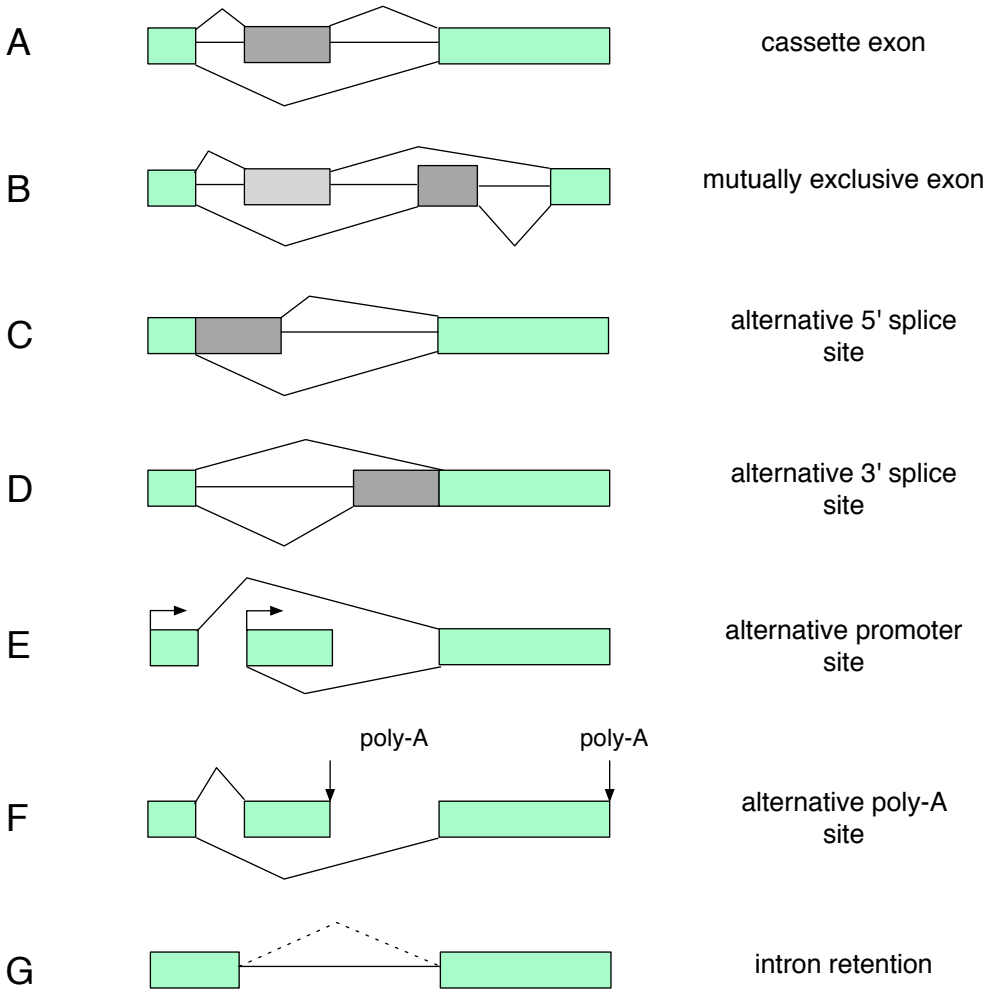

Figure 1-2: Seven basic AS patterns in multicellular eukaryotes.

Schematics of hypothetic pre-mRNAs are shown under different AS patterns, with green rectangles representing exons and gray rectangles represent alternatively spliced exons or exonic regions. Arrows show transcription start site. Arrowheads show poly-A sites.

In addition to serving as a major source of biological diversity, AS also plays important roles in regulatory networks. There are many mechanisms of AS in biological regulation. As one mechanism, AS enables the generation of functionally distinct protein

products from the same gene in response to different stimuli, and can work as a switch in cell fate determination (Figure 1-3A). For example, the gene encoding Bcl-2-like protein 1 (Bcl2l1 or Bcl-x), a member of the Bcl-2 family of apoptosis regulators, produces two major splicing isoforms—anti-apoptotic *Bclx-L* and pro-apoptotic *Bclx-S* due to the usage of an alternative 5' splice site in exon 2 (Figure 1-3A; Akgul et al. 2004). The switch of the two splicing isoforms of *Bcl-x* affects the cell's decision on whether to commit apoptosis. Besides *Bcl-x*, several other apoptosis regulators also generate pro- and anti-apoptosis isoforms (Akgul et al. 2004). As another mechanism, AS can introduce premature termination codons in an mRNA, leading to degradation of the mRNA through nonsense-mediated mRNA decay (AS-NMD) (Lewis et al. 2003) (Figure 1-3B). Here, AS directly regulates the total abundance of a target, rather than the expression of specific isoforms. As an example, the *Drosophila* sex determination master gene Sex-lethal (Sxl) can produce the male-specific isoform or female-specific isoform based on the inclusion or exclusion of exon 3, which contains a premature stop codon (Figure 1-3B; Baker 1989). AS-NMD causes the degradation of the male-specific isoform and lack of functional Sxl protein in male flies, leading to male-related phenotypes. On the other hand, the female-specific isoform lacking exon 3 is translated into functional Sxl proteins in female flies. Sxl is itself a splicing regulator, and alters the splicing of a cascade of downstream mRNA transcripts, resulting in female-related phenotypes (Baker 1989). The two example AS mechanisms above can be found in many biological regulatory networks, serving as crucial determinants of cell fate and identity.

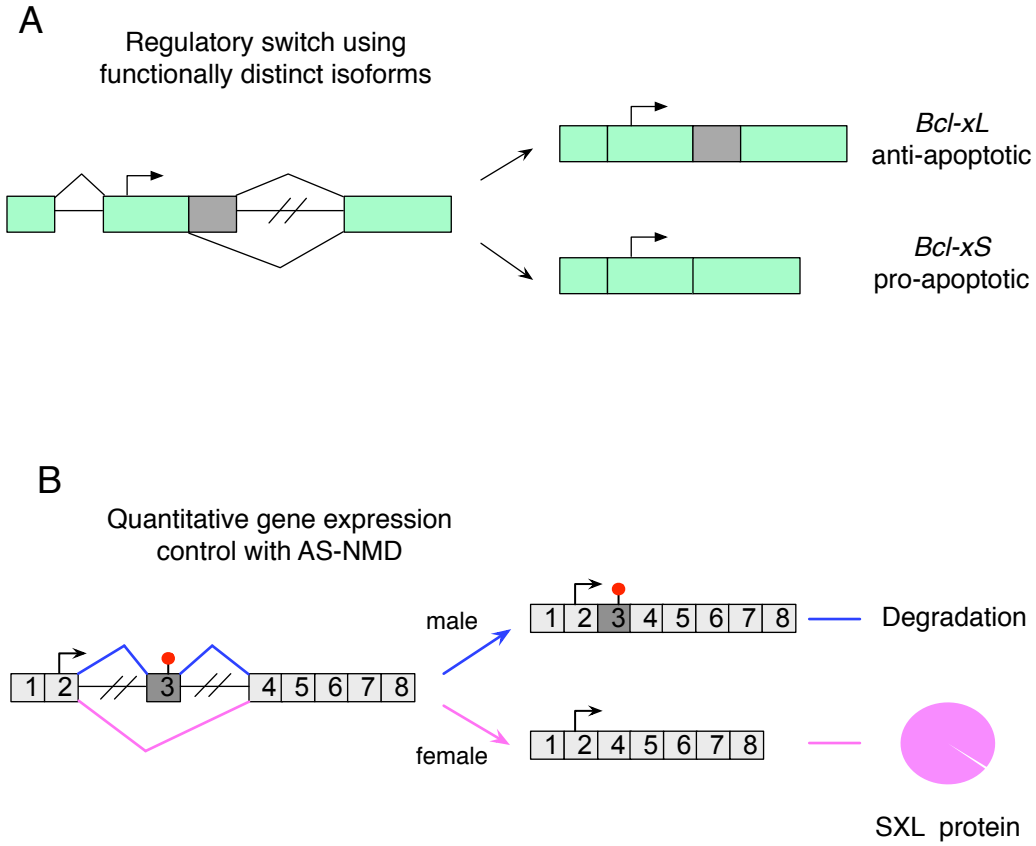

Figure 1-3: Two examples AS mechanisms in biological regulatory networks.

(A) AS enables *Bcl-x* to generate functionally distinct mRNA isoforms and protein variants to affect cells' decision in committing apoptosis. A schematic of *Bcl-x* is shown. Green rectangles represent exons and black lines between exons are introns. Gray rectangle represents the alternatively spliced exonic region. Arrows show translation start site.

(B) AS regulates the abundance of *Sxl* through AS-NMD in *Drosophila*. A schematic of *Sxl* is shown. Gray rectangles represent exons and black lines between exons are introns. Dark rectangle represents the alternatively spliced exon 3, with the stop sign representing a premature termination codon.

A HIERARCHICAL CONTROL OF ALTERNATIVE SPLICING REGULATION

The selection of splice sites by the spliceosome determines AS decisions. In yeast where constitutive splicing is dominant and AS is rare, the sequence elements for 5'SS, BPS and 3'SS are conserved and well recognized by U1 snRNP, SP1 and U2AF of the spliceosome machinery (Figure 1-1A; Izquierdo et al. 2006). However in higher eukaryotes where AS is more a rule than exception, the splice site sequence elements are degenerate and not strong enough to direct the spliceosome on their own (Figure 1-1A).

AS in multicellular eukaryotes is under the regulation of a large repertoire of *trans*-acting non-snRNP RNA-binding proteins (RBPs) and additional *cis*-elements in the pre-mRNA sequences (Wahl et al. 2009). Two classical classes of these RBPs are serine-arginine repeat proteins (SR proteins) and heterogeneous nuclear ribonucleoproteins (hnRNPs) (Matlin et al. 2005). RBPs recognize *cis*-elements within the pre-mRNA sequences and help to recruit or dissociate the spliceosome machinery to/away from a particular splice site (Figure 1-4A; Singh and Valcárcel 2005). These *cis*-elements associated with RBPs are short regulatory sequences flanking splice sites, including exonic or intronic splicing enhancers or silencers (Figure 1-4A). Collectively, AS decision is the combinatorial result from the 'splicing code' embedded in the pre-mRNA sequences and a large repertoire of proteins devoted to interpreting the code (Matlin et al. 2005).

The splicing code in AS is analogous to transcription factor binding sites in DNA sequences in the transcription network. Much effort has been dedicated to deciphering the splicing code (Wang and Burge 2008; Hartmann and Valcárcel 2009; Nilsen and Graveley 2010), although the process proves to be experimentally and computationally

Figure 1-4: A hierarchical control of AS regulation.

(A) The interplay of *cis*-elements within the pre-mRNA sequence and *trans*-acting RBPs in AS regulation. A schematic of a hypothetic pre-mRNA is shown. Green rectangles represent exons and black lines between exons are introns. 5' splice sites are marked by "GU", branch sites by A and 3' splice sites by "AG". Orange and red bands within the middle exon represent an exonic splicing enhancer (ESE) and an exonic splicing silencer (ESS). Blue short lines within introns represent an intronic splicing enhancers (ISE) and an intronic splicing silencers (ISS). SR proteins and hnRNPs can bind to these *cis*-elements and cast positive (+) or negative (-) effects on the selection of the nearby splice sites.

(B) A hierarchical control of AS regulation in multicellular eukaryotes. AS decision is under the regulation of *cis*-elements (red short lines) within pre-mRNA sequences and a large repertoire of RNA-binding proteins (RBPs, purple and blue circles). These RBPs are further controlled by many regulatory mechanisms that can cause tissue-specific synthesis, re-localization or modification of these RBPs so as to affect AS decisions. Circled P on one of the RBPs represents phosphorylation as an example of post-translational modification that can change the activities of RBPs. These regulatory mechanisms on RBPs are incorporated to the cellular signaling pathways and respond well to environmental cues.

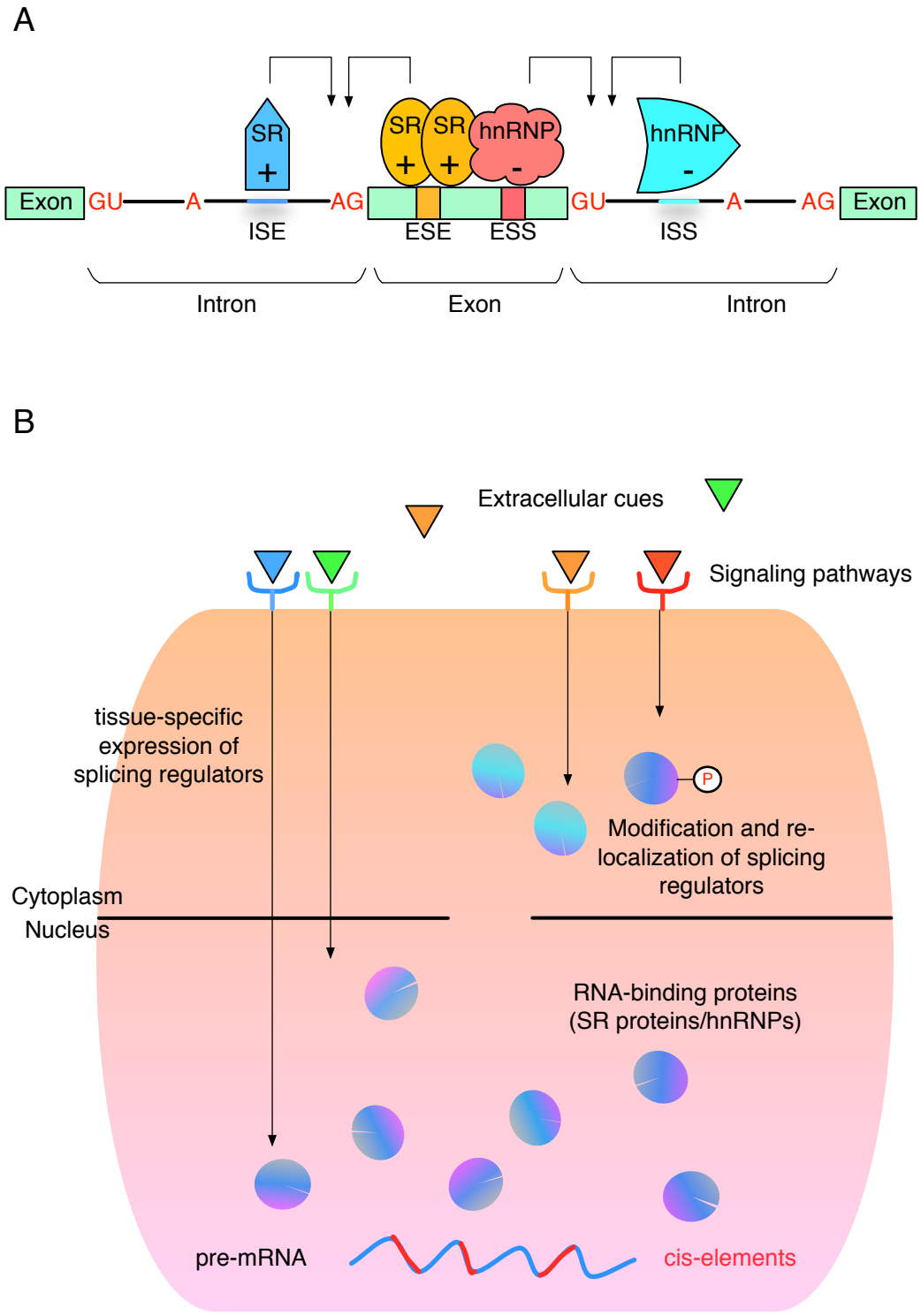

Figure 1-4 (Continued).

challenging for several reasons. *Cis*-element sequences in multicellular eukaryotes are usually short and degenerate, occurring frequently in the genome and not highly specific (Singh and Valcárcel 2005). In addition, the RBPs bind to *cis*-elements and associate with the spliceosome with low affinity. The decision on a given splice site is the combinatorial product of many antagonistic-acting splicing factors (Smith and Valcárcel 2000), complicating the prediction of AS decisions (Figure 1-4A). However, it is possible that examination of AS at a global scale with fast developing high-throughput techniques will improve our understanding of the combinatorial interplay between the *cis*-elements and the RBPs in AS decisions, and contribute to decipher the splicing code. Recently, some groups applied high-throughput genomic technologies to seek the splicing code from analyzing large samples of splicing events in different tissues and cell types (Barash et al. 2010). The resulted code is used to predict AS patterns under specific contexts and received positive results (Barash et al. 2010).

To further complicate the regulation of AS, there are many cellular regulatory mechanisms that regulate the activities of RBPs and consequently affect AS decisions (Moore and Silver 2008). These mechanisms span from transcriptional to post-translational levels, and can cause significant AS changes through modification, re-localization or tissue-specific synthesis of RBPs (Figure 1-4B; Moore and Silver 2008; Stamm 2002). For example, post-translational modifications such as protein phosphorylation can change the location and activities of RBPs and alter splice site choices (Stamm 2008). These regulatory mechanisms on RBPs are responsive to cellular signaling pathways and environment cues, establishing a hierarchical control of AS regulation (Figure 1-4B; Moore and Silver 2008).

ABERRANT SPLICING AND HUMAN DISEASES

AS is strictly regulated and provides diverse mRNA profiles in response to environmental cues. However, such diversity and flexibility come at a price: mis-regulation and defects can occur at multiple levels of AS regulation and cause aberrant splicing leading to human diseases. Indeed, it is estimated that 15% and maybe as many as 50% of disease-causing human hereditary mutations affect splicing decisions (Ward and Cooper 2010; López-Bigas et al. 2005). Aberrant AS is most frequently involved in diseases associated with functions of complex tissues or regulations of intricate biological processes (Blencowe 2006; Mills and Janitz 2012). As the most structurally complicated organ in animals, the brain generates more AS transcripts than all other tissues (Johnson et al. 2009; Yeo et al. 2004). It is perhaps not a surprise, then, that a fast growing list of neurological disorders are found associated with misregulaton in AS (Licatalosi and Darnell 2006). Apoptosis is a complicated cellular process under many layers of regulation, part of which is from AS (Akgul et al. 2004). To evade strict control of apoptosis, tumors show significantly altered AS profiles compared to normal tissues (Venables et al. 2009), establishing a critical role of AS misregulation in cancer.

Disease-related aberrant splicing can occur from either disruptions in *cis*-elements within a pre-mRNA sequence or from malfunctioning *trans*-acting RBPs, and can thus be classified into *cis*-acting and *trans*-acting splicing defects. Mutations in the core splicing machinery can also cause context-specific splicing deficiencies.

Cis-acting splicing defects: disruption of the 'splicing code'

Cis-acting splicing defects are caused by mutations in splice sites and auxiliary *cis*-elements in the pre-mRNA sequences, disrupting the splicing code. The effect of *cis*-acting splicing defect is usually local, affecting only specific splicing decisions on a particular pre-mRNA.

One example of *cis*-acting splicing defects is from the disease frontotemporal dementia with Parkinsonism linked to chromosome 17 (FTDP-17). This disease is caused by any of a series of mutations in the gene encoding microtubule-associated protein tau (MAPT) that is predominantly expressed in the central nervous system (Faustino and Cooper 2003). Exon 10 of MAPT encodes the last of four microtubule-binding domains in the protein and can be alternatively included or excluded, generating MAPT proteins with either three or four microtubule-binding domains (3R-tau or 4R-tau) (Goedert et al. 1989; Andreadis et al. 1992). FTDP-17 is associated with aggregations of MAPT in neuronal cytoplasmic inclusions, and 4R-tau is predominantly in MAPT aggregates in individuals with FTDP-17 (Buée et al. 2000; Hutton et al. 1998; Spillantini et al. 1998). The 4R-tau/3R-tau ratio of MAPT is strictly regulated by exonic and intronic splicing enhancers/silencers around exon 10 in the pre-mRNA sequence (Faustino and Cooper 2003). In normal brain tissues, MAPT produces equal amounts of 4R-tau and 3R-tau (Goedert and Jakes 1990; Kosik et al. 1989). One class of MAPT mutations in FTDP-17 changes the splicing *cis*-elements around exon 10 and increases the production of 4R-tau. As low as a two-fold change in the 4R/3R ratio could result in disease syndromes

(Faustino and Cooper 2003). Thus, the mis-splicing of exon 10 in MAPT due to mutations in the pre-mRNA *cis*-elements underlies the pathology of FTDP-17.

Trans-acting splicing defects: disruption of the RNA splicing 'regulon'

Trans-acting splicing defects are caused by malfunctioning splicing regulators like RBPs. In AS, the splicing of functionally related pre-mRNAs is usually coordinately regulated by one or a few splicing regulators. This phenomenon is proposed as RNA splicing 'regulon' (Keene 2007) that is analogous to operons in prokaryotes, where a cluster of genes are under the regulation of a single promoter (Jacob et al. 1960), or transcriptional regulation in eukaryotes where multiple functionally related genes are regulated by a single transcription factor. Consequently, *trans*-acting splicing defects are usually global and could affect the AS of many pre-mRNAs that are targets of the disrupted splicing regulator.

One example of *trans*-acting splicing defects is from the autoimmune disorder paraneoplastic opsoclonus-myoclonus ataxia (POMA). POMA is a paraneoplastic neurological syndrome in which neural damage is caused by remote effects of cancer unrelated to metastasis, tumor-related infection or metabolism disruption (Darnell and Posner 2003). Disruption of Nova-1 function is the primary cause of POMA. Nova-1 is a member of the neuro-oncological ventral antigen (Nova) family and is an RNA-binding protein exclusively expressed in the central nervous system (Buckanovich et al. 1993; Buckanovich et al. 1996). Nova-1 regulates the AS of a network of pre-mRNAs whose protein products are functionally enriched in neuron synapse formation and development (Ule et al. 2005). In POMA, tumors express antigens that are identical to Nova-1 and

cause immune responses toward the tumor as well as endogenous, Nova-1-expressing neurons (Luque et al. 1991; Darnell and Posner 2003). The autoimmune attack on Nova-1 inhibits Nova-1–RNA interactions and disrupts Nova-1's regulation of neuron-specific AS (Buckanovich et al. 1996; Jensen et al. 2000; Ule et al. 2005). The collective effects of splicing changes in the Nova-1–regulated mRNA network leads to neurodegenerative phenotypes observed in POMA (Licatalosi and Darnell 2010).

As another example, studies on cancer-associated AS provide evidence that aberrant splicing patterns in cancer are mostly caused by global splicing alteration due to the malfunctioning of *trans*-acting splicing factors (Wang and Cooper 2007; Ward and Cooper 2010). For instance, SR protein ASF/SF2 is recognized as a proto-oncogene and is over-expressed in many tumors (Karni et al. 2007). The oncogenic properties of ASF/SF2 can be directly linked to its splicing regulation of many downstream cancer-related genes (Wang and Cooper 2007; Karni et al. 2007).

Pathological mutations in the core splicing machinery

There have been very few cases where mutations in the core splicing machinery are found to be associated with human diseases. It is possible that defects in core spliceosome components are generally incompatible with the survival of organisms, whereas errors in specific AS decisions can be sufficiently tolerable to cause diseases but not immediate death. One surprising and poorly understood observation is that mutations in universally expressed core spliceosomal components can cause highly tissue-specific pathologies.

Retinitis pigmentosa (RP) is a degenerative eye disease characterized by a gradual vision loss that often leads to blindness, due to progressive death of rod photoreceptor cells of the retina. Many genetic mutations have been discovered to be associated with this disease and among them are HPRP3, PRP31, and PRPC8—factors crucial for the assembly and functioning of U4/U5/U6 tri-snRNPs in the spliceosome (McKie et al. 2001; Vithana et al. 2001; Chakarova et al. 2002). These discoveries indicate that certain defects in the spliceosome assembly pathway caused by the mutations may be behind the death of photoreceptor cells and the symptoms in RP. However, how mutations in the universal splicing machinery could lead to retina-specific pathological manifestations but not in other tissues remains unclear.

ALTERNATIVE SPLICING REGULATION AND HUMAN DISEASES: A SYSTEMS BIOLOGY APPROACH

The diverse splicing patterns and the hierarchical control of AS regulation make it necessary to investigate AS at a global level. Systems biology provides effective tools and perspectives for addressing such complex biological phenomena, using a holistic perspective (Alon 2006). Systems biology is an interdisciplinary field combining tools from mathematics, computer science, experimental biology and other subjects with an aim of understanding such complex biological systems.

High-throughput tools for systematic AS studies

Several genome-wide approaches have been developed to study AS from a systematic perspective. Different kinds of microarrays have been designed to investigate

global AS profiles in eukaryotes, such as exon-junction arrays (with probes spanning splicing junctions), exon tiling arrays (with probes within every known and predicted exons), and genomic tiling arrays (with probes tiled across the whole genome) (Moore and Silver 2008). More recently, whole transcriptome sequencing (RNA-seq) technology has provided an unbiased method for AS profiling by direct sequencing. RNA-seq avoids many drawbacks of microarrays, including the limitations from array probe design and biases from probe cross-hybridization, and is now the preferred method for global AS profiling. Microarrays and RNA-seq have both been applied to study the global AS profiles of normal or disease-associated tissues and cell types, and have been used to detect genome-wide AS changes upon the perturbation of a splicing regulator (Wang et al. 2008; Pan et al. 2008; Ule et al. 2005; Gehman et al. 2011; Charizanis et al. 2012).

The combination of high-throughput techniques and molecular experimental methods is greatly improving our understanding of AS regulation, especially the regulation of *trans*-acting splicing factors. Specifically, CLIP (cross-linking and immunoprecipitation) provides an effective way to study the target pre-mRNAs of a splicing factor (Ule et al. 2003). CLIP employs UV-crosslinking between the RNA-binding regulatory protein and its target RNAs followed by immunoprecipitation with antibodies against the protein (Ule et al. 2003). CLIP can be used in conjunction with high-throughput sequencing in a procedure called CLIP-seq: cross-linking immunoprecipitation-high-throughput sequencing. CLIP-seq has been used to identify the target mRNA network of splicing factors, and to characterize the specific *cis*-elements within the target mRNAs that these splicing factors bind to (Ule et al. 2003; Licatalosi et al. 2008).

Computational challenges in evaluating global alternative splicing

Next-generation sequencing allows us to study AS on a genome-wide scale. However, the rapid development of the sequencing technology itself brings forward a new challenge of analyzing the massive amount of data generated. For microarray data, numerous methods have been developed to quantify splicing variants and analyze differential AS between conditions (Moore and Silver 2008). RNA-seq technique is a more powerful technique, but it necessitates a new quantification paradigm by digitalizing the expression level of each mRNA transcript with read counts. Computational methods to quantify global AS and analyze differential AS using read counts from RNA-seq are still under intensive development (Mortazavi et al. 2008; Trapnell et al. 2010; Katz Y et al. 2010; Anders and Huber 2010). Moreover, the methods for AS quantification from RNA-seq data are also dependent on the ongoing development of basic RNA-seq analysis tools, such as mapping a tremendous amount of short reads to the reference genome and *ab initio* building of genomic structure from the sequencing data (Trapnell et al. 2009; Langmead et al. 2009; Trapnell et al. 2010; Guttman et al. 2010).

Screening for unidentified links between alternative splicing regulation and human diseases

High-throughput cell-based screening is a powerful strategy to identify genes or compounds that are involved in the regulation of a biological process (An and Tolliday 2010). The broad, unbiased scope of screening facilitates the recognition of previously

unidentified functions of genes, including disease associations. Several cellular screens have been performed to study the effect of potential factors or chemicals on specific AS events using splicing-sensitive reporters (Bonano et al. 2006; Orengo et al. 2006; Stoilov et al. 2008; Warzecha et al. 2009). These reporters are usually designed so that they produce different fluorescent or chemiluminscent products depending on how they are spliced (Cooper 2005). As an example, a MAPT exon 10-sensitive splicing reporter was built so that *RFP* is generated when exon 10 is included but *GFP* is produced if exon 10 is skipped (Stoilov et al. 2008). In a small-molecule screen using this system researchers successfully identified cardiotonic steroids as a regulator for the splicing of exon 10 in MAPT (Stoilov et al. 2008). Recently, the implementation of genome-wide RNA-interference (RNAi) screening has provided a way to examine the effects brought to the cell by loss-of-function of every known gene in an organism. Splicing sensitive reporters can then identify any potential factor whose loss-of-function affects the splicing event studied (Boutros et al. 2008; Oberdoerffer et al. 2008; Venables et al. 2008).

Dynamics of the alternative splicing regulatory network and applications in synthetic biological circuits

AS regulatory networks incorporate standard network motifs such as positive and negative feedback loops. For instance, Sxl, the master sex determinant in *Drosophila*, is a splicing factor and auto-regulates its own pre-mRNA splicing toward the female-specific isoform. This positive feedback loop of Sxl splicing regulation maintains a steady production of Sxl protein in female flies (Baker 1989; Figure 1-3B). A unique feature of the post-transcriptional regulation such as the AS regulatory network is that it enables

cells to change protein levels rapidly in response to environmental cues without transcript synthesis (Pleiss et al. 2007). This feature suggests the existence of kinetic mechanisms that link rapid regulatory response to the combinatorial binding of splicing factors to *cis*-elements and splicing activating by the spliceosome. Multiple kinetic models have been proposed for transcription factors binding DNA and RNA Polymerase II activating transcription, with model behaviors highly predictive of the in vivo experimental readout (Tay et al. 2010; Raj et al. 2006). It is of interest to model potential mechanisms of splicing factors binding pre-mRNA and triggering splicing, and compare the response time scale to a model of transcriptional regulation.

If indeed a unique, fast-responding kinetic mechanism of AS regulatory network is confirmed, the network motifs could have great potentials in synthetic biology. Synthetic biology is an emerging field that utilizes natural biological regulatory mechanisms to design and constructs synthetic circuits for industry, medicine and other application areas (Serrano 2007). For example, devices that can retain memory of transient exposures to a pulse signal have many potential applications in progressive diseases such as cancer. Such devices can track cells that have experienced a transient harmful stimulus and follow the subsequent cancerous development. A speedy memory response to the transient signal will be crucial in the performance of such devices. Synthetic memory circuits built to-date have mostly relied on the use of transcriptional regulatory network motifs (Vilaboa et al. 2005; Ajo-Franklin et al. 2007; Burrill and Silver 2011; Burrill et al. 2012). Especially in mammalian cells, these circuits show heterogeneous performances in retaining memory among cells (Burrill and Silver 2011; Burrill et al. 2012). This performance inconsistency is potentially due to the relatively

slow response time of the transcription-based device versus that of the cell cycle (Burrill and Silver 2011). In contrast, using a natural "memory device," every *Drosophila* cell homogeneously remembers its sex due to the splicing of *Sxl* through differentiation, development and many generations of cell divisions. AS regulatory network is currently underutilized in synthetic biology. Elucidation of AS regulatory network dynamics will greatly increase its potential applications in synthetic biological circuits.

THIS WORK

The past three decades' research in molecular biology, genetics and cell biology have established mechanisms of AS, the regulation of AS and the important roles AS plays in generating biological diversity, controlling regulatory pathways, and influencing human diseases. However, instead of getting a relatively complete atlas of knowledge on AS, we have discovered AS to be a far more complicated process than expected. Recent studies have investigated AS at a global level, and revealed AS to involve a complex, highly dynamic, and context-specific regulatory network. Consequently, new challenges have been brought forth in the field: to explore the significance and functions of the global AS network in different biological and physiological contexts will require new tools and a more integrated view of post-transcriptional regulation. The work presented here aims to address these challenges with a systems biology approach and examines the functions of AS regulatory network in a pathological context.

Chapter 2 describes work in studying AS regulation in apoptosis, a programmed cell death pathway that is closely associated with many important biological processes and diseases, specifically cancer. In this study, a genome-wide screen was carried out

through splicing-sensitive reporters of *Bclx* and *Mcl1*, both apoptosis regulators that can generate antagonistic splicing variants regulating apoptosis. The screen identified >150 novel factors that affect the AS of *Bclx* and *Mcl1* and are involved in apoptosis decisions through the modulation of anti-apoptotic and pro-apoptotic splicing variants. This screen revealed novel functional links between cell-cycle control and apoptosis through an AS network. Specifically, aurora kinase A is identified to be a potent factor that, when inhibited, can cause significant splicing shift towards pro-apoptotic variants in both *Bclx* and *Mcl1*. The inhibition of aurora kinase A can also cause AS changes in other apoptosis-related transcripts, including a pro-apoptotic shift in the splicing of *caspase-9*. These pro-apoptotic AS shifts in the transcripts were linked to the turnover of ASF/SF2, which directly binds the target pre-mRNAs and regulates their splicing. In addition, this screen unearthed several factors related to various diseases for which a link to aberrant splicing was not previously recognized.

Chapter 3 describes Polyglutamine-binding protein 1 (PQBP1)'s function in AS regulation and the effect of PQBP1-mediated AS regulatory network in neurons and neurological disorders. PQBP1, a factor associated with a series of X-linked mental retardation diseases, was unexpectedly identified as a splicing effector from the screen described in Chapter 2. We found that PQBP1 is capable of influencing the splicing of multiple mRNAs and is associated with many splicing factors, including the key U2 snRNP component SF3B1. Loss of functional PQBP1 reduced SF3B1-substrate mRNA association and caused significant changes in AS decisions. Depletion of PQBP1 in mouse primary neurons led to defects in neurite outgrowth and altered the AS of a network of pre-mRNAs that encode factors functionally enriched in neuron projection

development. Interestingly, disease mutants of PQBP1 lose associations with splicing factors and cannot complement the neurite outgrowth defects and aberrant AS patterns found in PQBP1-depleted cells. Collectively, this work shows that PQBP1 affects AS of many pre-mRNAs and that aberrant splicing caused by malfunctioning of PQBP1 may contribute to the pathology of PQBP1-linked neurological disorders.

Chapter 4 presents a junction-based method for differential analysis of global AS under different conditions using RNA-seq data. This method is among the first explorations to quantify AS using only read counts that are mapped directly to splicing junctions. Statistical methods were developed to perform differential analysis on the 'usage' (number of mapped reads) of competitive AS junctions and decide if significant AS changes occur among the junctions between conditions. This method does not depend on prior knowledge of the genome annotation and performs a thorough and unbiased investigation of AS patterns under different conditions. Moreover, this method integrates variability among biological replicates, producing reliable statistical tests of global AS changes. Using this method we identified 585 AS events that experience significant changes upon PQBP1 knockdown in mouse primary cortical neurons, distributed among 457 genes. Ten identified potential PQBP1 AS targets were randomly selected for experimental validation and all were verified by RT-PCR. One of the experimentally validated targets, *Ncam1* can only be identified as an AS target of PQBP1 using our method but not other currently used bioinformatics tools, due to an extremely short alternatively spliced exon in *Ncam1*. These results demonstrated the specificity and sensitivity of our method.

Chapter 5 presents a synthetic circuit in mammalian cells that confers memory of extracellular stimuli based on the auto-regulatory feedback loop from the splicing regulation of Sex-Lethal (Sxl), the *Drosophila* sex determination master gene. We show that the positive feedback loop of Sxl regulating the splicing of its own pre-mRNA can be transported from *Drosophila* to mammalian cells. Furthermore, with quantitative modeling and experimental data, we show that this positive feedback loop can serve as a device to rapidly establish memory of transient exposure to doxycycline stimuli in mammalian cells. This memory device can be applied to the study of disease development in progressive disorders, such as investigating the long-term effects of transient hypoxia stimuli in early cancerous cells.

REFERENCE

Ajo-Franklin CM, Drubin DA, Eskin JA, Gee EP, Landgraf D, Phillips I, Silver PA. 2007. Rational design of memory in eukaryotic cells. *Genes Dev* **21:** 2271-2276.

Akgul C, Moulding DA, Edwards SW. 2004. Alternative splicing of Bcl-2-related genes: functional consequences and potential therapeutic applications. *Cell Mol Life Sci* **61:** 2189–2199.

Alon, U. 2006. *An introduction to Systems Biology: design principles of biological circuits*. CRC Press, Boca Raton, FL.

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol* **11:** R106.

An WF, Tolliday N. 2010. Cell-based assays for high-throughput screening. *Mol Biotechnol* **45:** 180-186.

Andreadis A, Brown WM, Kosik KS. 1992. Structure and novel exons of the human tau gene. *Biochemistry* **31:** 10626-10633.

Baker BS. 1989. Sex in flies: the splice of life. *Nature* **340:** 521-524.

Barash Y, Calarco JA, Gao W, Pan Q, Wang X, Shai O, Blencowe BJ, Frey BJ. 2010. Deciphering the splicing code. *Nature* **465:** 53-59.

Berget SM, Moore C, Sharp PA. 1977. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc Natl Acad Sci U S A* **74:** 3171-3175.

Black DL. 2003. Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* **72**: 291–336.

Blencowe BJ. 2006. Alternative splicing: new insights from global analyses. *Cell* **126:** 37-47.

Bonano VI, Oltean S, Brazas RM, Garcia-Blanco MA. 2006. Imaging the alternative silencing of FGFR2 exon IIIb in vivo. *RNA* **12:** 2073-2079.

Boutros M, Ahringer J. 2008. The art and design of genetic screens: RNA interference. *Nat Rev Genet* **9:** 554-566.

Buckanovich RJ, Posner JB and Darnell RB. 1993. Nova, the paraneoplastic Ri antigen, is homologous to an RNA-binding protein and is specifically expressed in the developing motor system. *Neuron* **11:** 657-672.

Buckanovich RJ, Yang YY, Darnell RB. 1996. The onconeural antigen Nova-1 is a neuron-specific RNA-binding protein, the activity of which is inhibited by paraneoplastic antibodies. *J Neurosci* **16:** 1114-1122.

Buée L, Bussière T, Buée-Scherrer V, Delacourte A, Hof PR. 2000. Tau protein isoforms, phosphorylation and role in neurodegenerative disorders. *Brain Res Brain Res Rev* **33:** 95-130.

Burrill DR, Inniss MC, Boyle PM, Silver PA. 2012. Synthetic memory circuits for tracking human cell fate. *Genes Dev* **26:** 1486-1497.

Burrill DR, Silver PA. 2011. Synthetic circuit identifies sub-populations with sustained memory of DNA damage. *Genes Dev* **25:** 434-439.

Chakarova CF, Hims MM, Bolz H, Abu-Safieh L, Patel RJ, Papaioannou MG, Inglehearn CF, Keen TJ, Willis C, Moore AT, et al. 2002. Mutations in HPRP3, a third member of pre-mRNA splicing factor genes, implicated in autosomal dominant retinitis pigmentosa. *Hum Mol Genet* **11:** 87-92.

Charizanis K. Lee KY, Batra R, Goodwin M, Zhang C, Yuan Y, Shiue L, Cline M, Scotti MM, Xia G, et al. 2012. Muscleblind-like 2-mediated alternative splicing in the developing brain and dysregulation in myotonic dystrophy. *Neuron* **75:** 437-450.

Cooper TA. 2005. Use of minigene systems to dissect alternative splicing elements. *Methods* **37:** 331-340.

Cooper TA, Wan L, Dreyfuss G. 2009. RNA and disease. *Cell* **136:** 777-793.

Crick F. 1970. Central dogma of molecular biology. *Nature* **227:** 561-563.

Darnell RB, Posner JB. 2003. Paraneoplastic syndromes involving the nervous system. *N Engl J Med* **349:** 1543-1454.

Faustino NA, Cooper TA. 2003. Pre-mRNA splicing and human disease. *Genes Dev* **17:** 419-437.

Gehman LT, Stoilov P, Maguire J, Damianov A, Lin CH, Shiue L, Ares M Jr., Mody I, Black DL. 2011. The splicing regulator Rbfox1 (A2BP1) controls neuronal excitation in the mammalian brain. *Nat Genet* **43:** 706-711.

Gilbert W. 1978. Why genes in pieces. *Nature* **271:** 501.

Goedert M, Spillantini MG, Potier MC, Ulrich J, Crowther RA. 1989. Cloning and sequencing of the cDNA encoding an isoform of microtubule-associated protein tau containing four tandem repeats: differential expression of tau protein mRNAs in human brain. *EMBO J* **8:** 393-399.

Goedert M, Jakes R. 1990. Expression of separate isoforms of human tau protein: correlation with the tau pattern in brain and effects on tubulin polymerization. *EMBO J* **9:** 4225-4230.

Guttman M, Garber M, Levin JZ, Donaghey J, Robinson J, Adiconis X, Fan L, Koziol MJ, Gnirke A, Nusbaum C. et al. 2010. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat Biotechnol* **28:** 503-510.

Hartmann B, Valcárcel J. 2009. Decrypting the genome's alternative messages. *Curr Opin Cell Biol* **21:** 377-386.

Hutton M, Lendon CL, Rizzu P, Baker M, Froelich S, Houlden H, Pickering-Brown S, Chakraverty S, Isaacs A, Grover A. et al. 1998. Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17. *Nature* **393:** 702-705.

Izquierdo JM, Valcárcel J. 2006. A simple principle to explain the evolution of pre-mRNA splicing. *Genes Dev* **20:** 1679-1684.

Jacob F, Perrin D, Sanchez C, Monod J. 1960. Operon: a group of genes with the expression coordinated by an operator. *C R Hebd Seances Acad Sci* **250:** 1727-1729.

Jensen KB, Dredge BK, Stefani G, Zhong R, Buckanovich RJ, Okano HJ, Yang YY, Darnell RB. 2000. Nova-1 regulates neuron-specific alternative splicing and is essential for neuronal viability. *Neuron* **25:** 359-371.

Johnson MB, Kawasawa YI, Mason CE, Krsnik Z, Coppola G, Bogdanović D, Geschwind DH, Mane SM, State MW, Sestan N. 2009. Functional and evolutionary insights into human brain development through global Transcriptome analysis. *Neuron* **62:** 494-509.

Karni R, de Stanchina E, Lowe SW, Sinha R, Mu D, Krainer AR. 2007. The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nat Struct Mol Biol* **14:** 185-193.

Katz Y, Wang ET, Airoldi EM, Burge CB. 2010. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat Methods* **7:** 1009-1015.

Keene JD. 2007. RNA regulons: coordination of post-transcriptional events. *Nat Rev Genet* **8:** 533-543.

Keene JD, Lager PJ. 2005. Post-transcriptional operons and regulons co-ordinating gene expression. *Chromosome Res* **13:** 327-337.

Kosik KS, Orecchio LD, Bakalis S, Neve RL. 1989. Developmentally regulated expression of specific tau sequences. *Neuron* **2:** 1389-1397.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10:** R25.

Lewis BP, Green RE, Brenner SE. 2003. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc Natl Acad Sci USA* **100:** 189-192.

Licatalosi DD, Darnell RB. 2006. Splicing regulation in neurologic disease. *Neuron* **52:** 93-101.

Licatalosi DD, Mele A, Fak JJ, Ule J, Kayikci M, Chi SW, Clark TA, Schweitzer AC, Blume JE, Wang X et al. 2008. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456:** 464-469.

Licatalosi DD, Darnell RB. 2010. RNA processing and its regulation: global insights into biological networks. *Nat Rev Genet* **11:** 75-87.

Lodish H, Berk A, Kaiser CA, Krieger M, Scott MP, Bretscher A, Ploegh H, Matsudaira P. 2008. *Molecular cell biology*. W. H. Freeman and Company, New York, NY

López-Bigas N, Audit B, Ouzounis C, Parra G, Guigó R. 2005. Are splicing mutations the most frequent cause of hereditary disease? *FEBS Lett* **579:** 1900-1903.

Luque FA, Furneaux HM, Ferziger R, Rosenblum MK, Wray SH, Schold SC Jr, Glantz MJ, Jaeckle KA, Biran H, Lesser M, et al. 1991. Anti-Ri: an antibody associated with paraneoplastic opsoclonus and breast cancer. *Ann Neurol* **29:** 241-251.

McKie AB, McHale JC, Keen TJ, Tarttelin EE, Goliath R, van Lith-Verhoeven JJ, Greenberg J, Ramesar RS, Hoyng CB, Cremers FP, et al. 2001. Mutations in the pre-mRNA splicing factor gene PRPC8 in autosomal dominant retinitis pigmentosa (RP13). *Hum Mol Genet* **10:** 1555-1562.

Mills JD, Janitz M. 2012. Alternative splicing of mRNA in the molecular pathology of neurodegenerative diseases. *Neurobiol Aging* **33:** 1012.e11-e24.

Moore, MJ, Silver PA. 2008. Global analysis of mRNA splicing. *RNA* **14:** 197-203.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5:** 621–628.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409:** 860-921.

Mata J, Marguerat S, Bähler J. 2005. Post-transcriptional control of gene expression: a genome-wide perspective. *Trends Biochem Sci* **30:** 506-514.

Matlin AJ, Clark F, Smith CW. 2005. Understanding alternative splicing: towards a cellular code. *Nat Rev Mol Cell Biol* **6:** 286-398.

Nelson PT, Keller JN. 2007. RNA in brain disease: no longer just "the messenger in the middle". *J Neuropathol Exp Neurol* **66:** 461-468.

Nilsen TW, Graveley BR. 2010. Expansion of the eukaryotic proteome by alternative splicing. *Nature* **463:** 457-463.

Oberdoerffer S, Moita LF, Neems D, Freitas RP, Hacohen N, Rao A. 2008. Regulation of CD45 alternative splicing by heterogeneous ribonucleoprotein, hnRNPLL. *Science* **321:** 686-691.

Olson MV, Varki A. 2004. Genomics. The chimpanzee genome—a bittersweet celebration. *Science* **305:** 191-192.

Orengo JP, Bundman D, Cooper TA. 2006. A bichromatic fluorescent reporter for cell-based screens of alternative splicing. *Nucleic Acids Res* **34:** e148.

Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* **40:** 1413-1415.

Pleiss JA, Whitworth GB, Bergkessel M, Guthrie C. 2007. Rapid, transcript-specific changes in splicing in response to environmental stress. *Mol Cell* **27:** 928-937.

Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S. 2006. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol* **4:** e309

Rappsilber J, Ryder U, Lamond AI, Mann M. 2002. Large-scale proteomic analysis of the human spliceosome. *Genome Res* **12:** 1231-1245.

Schmucker D, Clemens JC, Shu H, Worby CA, Xiao J, Muda M, Dixon JE, Zipursky SL. 2000. Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* **101:** 671-684.

Serrano L. 2007. Synthetic biology: promises and challenges. *Mol Syst Biol* **3:** 158.

Singh R, Valcárcel J. 2005. Building specificity with nonspecific RNA-binding proteins. *Nat Struct Mol Biol* **12:** 645-653.

Smith CW, Valcárcel J. 2000. Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem Sci* **25:** 381-388.

Spillantini MG, Murrell JR, Goedert M, Farlow MR, Klug A, Ghetti B. 1998. Mutation in the tau gene in familial multiple system tauopathy with presenile dementia. *Proc Natl Acad Sci U S A* **95:** 7737-7741.

Stamm S. 2002. Signals and their transduction pathways regulating alternative splicing: a new dimension of the human genome. *Hum Mol Genet* **11:** 2409-2416.

Stamm S. 2008. Regulation of alternative splicing by reversible protein phosphorylation. *J Biol Chem* **283:** 1223-1227.

Stoilov P, Lin CH, Damoiseaux R, Nikolic J, Black DL. 2008. A high-throughput screening strategy identifies cardiotonic steroids as alternative splicing modulators. *Proc Natl Acad Sci U S A* **105:** 11218-11223.

Tay S, Hughey JJ, Lee TK, Lipniacki T, Quake SR, Covert MW. 2010. Single-cell NF-kappaB dynamics reveal digital activation and analogue information processing. *Nature* **466:** 267-271.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-seq. *Bioinformatics* **25:** 1105-1111.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28:** 511-515.

Ule J, Jensen KB, Ruggiu M, Mele A, Ule A, Darnell RB. 2003. CLIP identifies Nova-regulated RNA networks in the brain. *Science* **302:** 1212–1215.

Ule J, Ule A, Spencer J, Williams A, Hu JS, Cline M, Wang H, Clark T, Fraser C, Ruggiu M, et al. 2005. Nova regulates brain-specific splicing to shape the synapse. *Nat Genet* **37:** 844-852.

Venables JP, Koh CS, Froehlich U, Lapointe E, Couture S, Inkel L, Bramard A, Paquet ER, Watier V, Durand M et al. 2008. Multiple and specific mRNA processing targets for the major human hnRNP proteins. *Mol Cell Biol* **28:** 6033-6043.

Venables JP, Klinck R, Koh C, Gervais-Bird J, Bramard A, Inkel L, Durand M, Couture S, Froehlich U, Lapointe E, et al. 2009. Cancer-associated regulation of alternative splicing. *Nat Struct Mol Biol* **16:** 670-676.

Vilaboa N, Fenna M, Munson, J, Roberts, SM, Voellmy, R. 2005. Novel gene switches for targeted and timed expression of proteins of interest. *Mol Ther* **12:** 290-298.

Vithana EN, Abu-Safieh L, Allen MJ, Carey A, Papaioannou M, Chakarova C, Al-Maghtheh M, Ebenezer N.D, Willis C, Moore AT, et al. 2001. A human homolog of yeast pre-mRNA splicing gene, PRP31, underlies autosomal dominant retinitis pigmentosa on chromosome 19q13.4 (RP11). *Mol. Cell* **8:** 375-381.

Wahl MC, Will CL, Lührmann R. 2009. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136:** 701–718.

Wang GS, Cooper TA. 2007. Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat Rev Genet* **8:** 749-761.

Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB. 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature* **456:** 470–476.

Wang Z, Burge CB. 2008. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA* **14:** 802-813.

Ward AJ, Cooper TA. 2010. The pathobiology of splicing. *J Pathol* **220:** 152-163.

Warzecha CC, Sato TK, Nabet B, Hogenesch JB, Carstens RP. 2009. ESRP1 and ESRP2 are epithelial cell-type-specific regulators of FGFR2 splicing. *Mol Cell* **33:** 591-601.

Yeo G, Holste D, Kreiman G, Burge CB. 2004. Variation in alternative splicing across human tissues. *Genome Biol* **5:** R74.

CHAPTER 2: AN ALTERNATIVE SPLICING NETWORK LINKS CELL-CYCLE CONTROL TO APOPTOSIS

ABSTRACT

Alternative splicing is a vast source of biological regulation and diversity that is misregulated in cancer and other diseases. To investigate global control of alternative splicing in human cells, we analyzed splicing of mRNAs encoding Bcl2 family apoptosis factors in a genome-wide siRNA screen. The screen identified many regulators of *Bcl-x* and *Mcl1* splicing, notably an extensive network of cell-cycle factors linked to aurora kinase A. Drugs or siRNAs that induce mitotic arrest promote proapoptotic splicing of *Bcl-x*, *Mcl1*, and *caspase-9* and alter splicing of other apoptotic transcripts. This response precedes mitotic arrest, indicating coordinated upregulation of prodeath splice variants that promotes apoptosis in arrested cells. These shifts correspond to posttranslational turnover of splicing regulator ASF/SF2, which directly binds and regulates these target mRNAs and globally regulates apoptosis. Broadly, our results reveal an alternative splicing network linking cell-cycle control to apoptosis.

INTRODUCTION

Nearly all human precursor messenger RNAs (pre-mRNAs) undergo alternative splicing (AS), with tremendous variation and specificity across tissues, development, and disease (Pan et al. 2008). This vast complexity is a formidable challenge to experimental and informatic analysis of AS and its physiological roles (Blencowe 2006). One emergent concept is that functionally coherent transcript populations, termed RNA "regulons," are coregulated by dedicated RNA-binding proteins (RBPs) to promote specific biological functions (Keene 2007). These RBPs, notably SR proteins and hnRNPs, "decode" transcript cis elements and control stepwise assembly of spliceosomal snRNPs: U1 at the 5' splice -site (ss), then U2 near the 3' ss, and finally U5/4/6 (Wahl et al. 2009; Black 2003). Splicing control is integrated with signal transduction pathways, promoting dynamic, context-driven regulation. Collectively, these features produce robust cell- and tissue-specific signatures of exon use that shape many aspects of cell fate (Moore and Silver 2008). Exon signatures are radically transformed in tumors, but the causes and consequences are unknown.

In this study, we examine AS of Bcl2 family apoptosis regulators *Bcl-x* and *Mcl1*. Bcl2-like proteins contain up to four Bcl2-homology (BH) domains (BH1–4) (Hardwick and Youle 2009). Factors possessing all four BH domains, including Bcl2, Bcl-xL, and Mcl1L, antagonize apoptosis by preventing mitochondrial outer membrane permeablization (MOMP), thus sequestering proapoptotic factors in mitochondria. Factors lacking one or more BH domain, including Bid, BAD, and BAX, are proapoptotic and promote MOMP. A finely tuned balance of pro- and antiapoptotic Bcl2-like factors therefore controls mitochondrial integrity and hence downstream steps in

apoptosis such as apoptosome formation and caspase activation (Wang and Youle 2009). Remarkably, *Bcl-x*, *Mcl1*, and some other *Bcl2* family mRNAs are alternatively spliced to yield both long (L) antiapoptotic and short (S) proapoptotic forms. For *Bcl-x*, use of an alternative 5' ss in exon 2 excludes the BH1 and BH2 domains (Akgul et al. 2004). For *Mcl1*, exon 2 skipping excludes the BH1 and BH2 domains and eliminates the downstream transmembrane domain via frame shift.

Many *cis*-regulatory elements and *trans*-acting factors exert combinatorial control of *Bcl-x* splicing. Most known regulators, including Sam68, ASF/SF2, hnRNP F/H, SRp30c, and RBM25, altered *Bcl-x* AS in vitro or when overexpressed in cell culture (Cloutier et al. 2008; Zhou et al. 2008; Paronetto et al. 2007; Garneau et al. 2005). In addition, in RNA interference (RNAi)-based loss-of-function assays, depletion of Sam68 and hnRNPA1 favored *Bcl-xL* formation, while depletion of U2 snRNP component SF3B1/SAP155 favored *Bcl-xS* (Paronetto et al. 2007; Massiello et al. 2006). Comparably little is known of *Mcl1* splicing regulation.

Beyond Bcl2-like factors, caspases, "death receptors," ligands and various adaptors are regulated by AS, suggesting broad roles in controlling apoptosis (Schwerk and Schulze-Osthoff, 2005). Many apoptosis regulators, including Bcl2-like proteins, are proto-oncogenes that contribute to apoptosis resistance in cancer (Letai 2008; Fesik 2005). Modulation of apoptotic factors by targeting the splicing machinery is thus an attractive strategy to facilitate tumor cell death. Furthermore, while the divergent functions of *Bcl-x* and *Mcl1* isoforms in apoptosis are well established, the physiological contexts and upstream regulation of their expression are poorly defined. These unanswered questions illustrate a pervasive challenge in defining physiological contexts

of AS regulation, because strategies for systematic evaluation of upstream regulation are limited.

Genome-scale screening of RNA regulatory events is complicated by the difficulty of visualizing RNAs in vivo, and the technical infeasibility of high-throughput measurements by RT-PCR and other methods. Splicing-sensitive fluorescent reporters are an alternative strategy that produce a robust, visual output suitable for screening efforts (Stoilov et al. 2008; Warzecha et al. 2009; Orengo et al. 2006). Here, we present high-throughput assays that recapitulate physiological regulation of *Bcl-x* and *Mcl1* AS. In a whole-genome siRNA screen, we identified new factors that regulate the balance of anti- and proapoptotic splice isoforms, with striking enrichment for cell-cycle factors. These results define functional interactions between the cell cycle and splicing machineries in human cells that manifest in a coordinated program of AS controlling apoptosis.

RESULTS

Reporter assays for Bcl-x and Mcl1 alternative splicing

To develop splicing assays for high-throughput analysis, we designed splicing-sensitive reporters for the Bcl2 family apoptosis regulators *Bcl-x* and *Mcl1*. 5' untranslated region (UTR), open reading frame (ORF), and intervening intron sequences for *Bcl-x* and *Mcl1* were cloned in C-terminal fusions with *Venus* (yellow) and *mCherry* (red) complementary DNAs (cDNAs), respectively (Figure 2-1A). In HeLa cells, these constructs expressed long and short spliced mRNAs at ratios similar to endogenous mRNAs (Figure 2-1B). To render constructs splicing-sensitive, premature termination codons (PTCs) were introduced in alternative exon regions exclusive to long splice

Figure 2-1: *Bcl-x* and *Mcl1* alternative splicing reporters.

(A) *Bcl-x* and *Mcl1* minigene splicing reporters are schematized. Green rectangles are exons, with dark gray as alternative regions. Black lines depict introns. Large arrows denote translation start sites, and arrowheads show primer sets used for qPCR. Inserted PTCs, exclusive to long (L) isoforms, are shown by red "stop signs."

(B) Minigene constructs are spliced like endogenous mRNAs. Splice isoforms from minigene- or mock-transfected (endogenous) HeLa cells were analyzed by RT-PCR with primer sets from (A). Because of high transient minigene expression, cDNA from minigene samples was diluted $1{:}10^5$ relative to mock (endogenous). At this dilution, endogenous mRNAs do not contribute to products in the "minigene" lanes.

(C) A PTC introduced to the alternative exon region eliminated Bcl-xL-Venus but not Bcl-xS-Venus protein expression. HeLa cells were transfected with *Bcl-x-Venus* (PTC-) or *Bcl-x-Venus-PTC* (PTC+) constructs, and lysates were analyzed by western with indicated antibodies. Migration of isoforms is indicated to the left of images (last two lanes overexposed to show Bcl-xS-Venus).

(D) SF3B1 knockdown activated proapoptotic splicing of *Bcl-x* and *Mcl1* reporters. Stable reporter lines were transfected with SF3B1 or non-targeting siRNA pools and imaged in the indicated channels. SF3B1 knockdown increased minigene expression versus controls in both cases. SF3B1 depletion also decreased CFP modestly, possibly reflecting a role in splicing the EF1α intron. Subsequent experiments showed this effect on CFP expression was restricted to SF3B1 (see Supplemental Figures S2-2B and S2-2F).

(E) Upregulation of *Bcl-xS* by SF3B1 depletion was confirmed by western analysis. Lysates from cells treated as in (D) were analyzed with antibodies indicated to the left of images. SF3B1 knockdown depleted SF3B1 and increased Bcl-xS-Venus relative to controls.

(F) Results in (D) and (E) were verified at the RNA level by RT-PCR. Migration of *Bcl-xS*, *Bcl-xL* (PTC causes nonsense-mediated decay), and *GAPDH* (loading control) are indicated to left of the images.

(G) Results in (D), (E), and (F) model the regulation of endogenous *Bcl-x* and *Mcl1* mRNAs. Unmodified HeLa cells were transfected with SF3B1 or control siRNAs. RT-PCR analysis confirmed that SF3B1 knockdown favored short, proapoptotic forms. *S*/*L* ratios are quantified to the right of the images; values are the means of three independent measurements ± SD, normalized to controls.

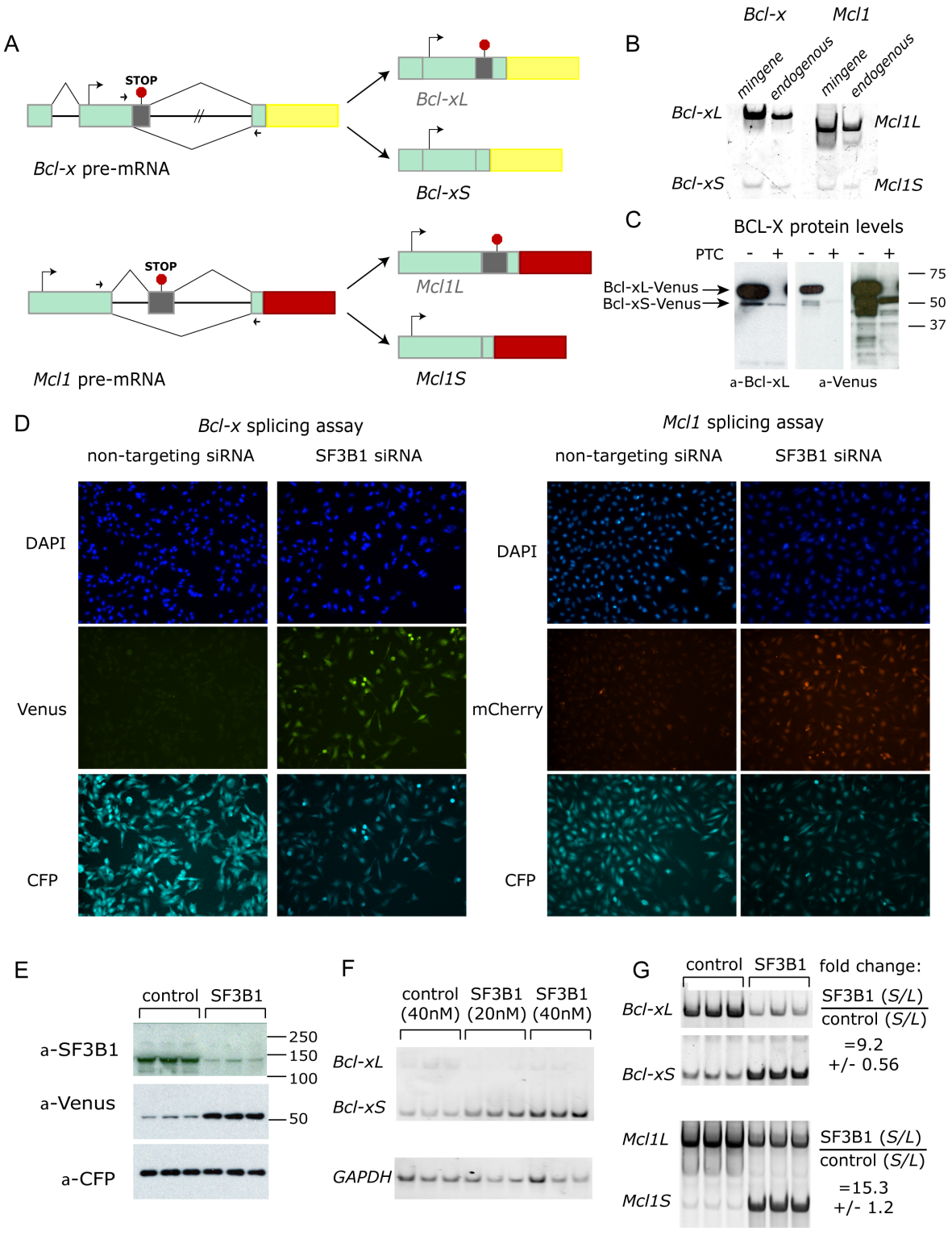

Figure 2-1 (Continued).

forms. As expected, PTCs eliminated expression of long protein variants, but short forms were retained (Figure 2-1C).

To produce screen assay cell lines, splicing reporters were stably transfected into HeLa cells along with a constitutive mCerulean fluorescent protein (CFP) construct. All constructs used the human EF1α promoter, which contains a 5' UTR intron, allowing dual measurements of minigene splicing and a constitutively spliced CFP reporter under identical control. To test the Bcl-x reporter line, we verified that siRNA-depletion of known regulator SF3B1 increased Bcl-xS-Venus expression relative to a non-targeting control (Figure 2-1D, left panels) (Massiello et al. 2006). Immunoblotting confirmed efficient siRNA knockdown of SF3B1, and upregulation of Bcl-xS-Venus reporter protein (Figure 2-1E). RT-PCR confirmed upregulation of the *Bcl-xS-Venus* mRNA (Figure 2-1F). Finally, RT-PCR analysis of endogenous *Bcl-x* transcript in HeLa cells verified that SF3B1 knockdown shifted splicing toward *Bcl-xS*, demonstrating congruous regulation of minigene and endogenous splicing (Figure 2-1G, upper panel).

We had no a priori knowledge of Mcl1 regulators, but SF3B1 knockdown also upregulated Mcl1S-mCherry in the splicing assay, establishing a positive assay control (Figure 2-1D, right panels). Analysis of endogenous *Mcl1* verified this shift toward *Mcl1S* (Figure 2-1G, lower panel).

High-throughput siRNA screens for alternative splicing regulators

To identify regulators of Bcl-x AS, >21,000 siRNA pools targeting known and predicted human genes were screened for upregulation of the *Bcl-x* reporter (Figure 2-2A). Three hundred and sixty-nine positive hits were identified with a Support Vector

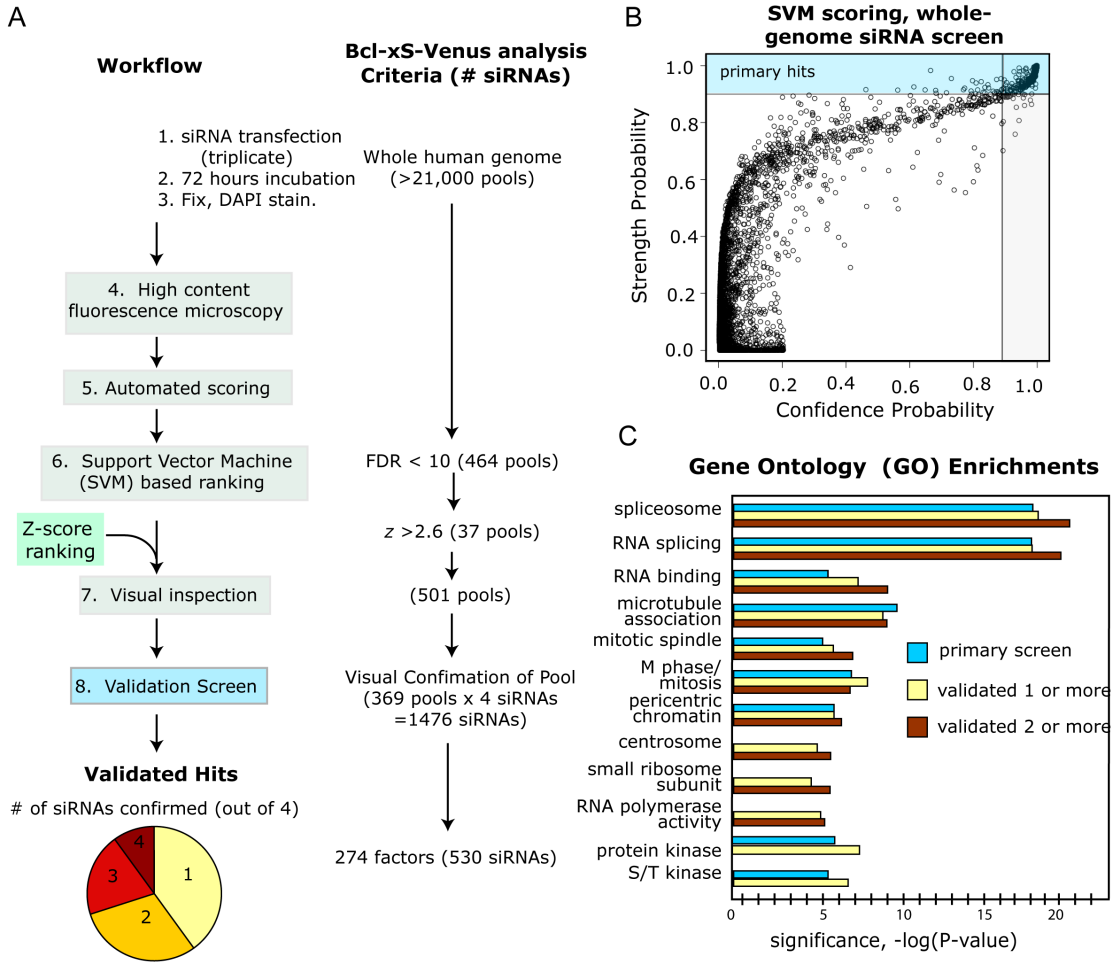

Figure 2-2: Whole-genome siRNA screen for regulators of *Bcl-x* alternative splicing.

(A) Screening, analysis, and hit selection are schematized. Of 369 primary hits, 274 revalidated. The pie chart indicates how many of 274 validated hits reconfirmed with one, two, three, or four siRNAs. One hundred and sixty "high-confidence" hits validated with two or more siRNAs (see the Supplemental Experimental Procedures for complete analysis methods).

(B) SVM probability scores are plotted for the whole-genome siRNA screen, with confidence on the x axis and strength on the y axis. The blue region shows the SVM cutoff for hit selection (see Figures S2-1, S2-2, and S2-3 and Tables S2-1 and S2-2 for detailed analysis and validation).

(C) GO functional enrichments from primary, revalidated (one or more siRNAs), and high-confidence (two or more siRNAs) hits are shown, plotted by relative statistical significance. See also Figures S2-1, S2-2, and S2-3 and Tables S2-1 and S2-2.

Machine (SVM) model that determined reproducibility (i.e., "confidence") across

triplicates and signal "strength" relative to positive and negative control siRNAs (Figure 2-2B, and Supplemental Figure S2-1 and Table S2-1). Hits had strong gene ontology (GO) enrichments for mRNA splicing/processing, protein kinase signaling, cytoskeleton association, and cell-cycle functions (Figure 2-2C). Importantly, the screen blindly recovered positive control SF3B1 and several of its interactors.

For validation, hits were retested in the screen assay with four individual siRNAs from deconvoluted SMARTpools. Two hundred and seventy-four of 369 factors validated with at least one siRNA, and 160 validated with two or more (Figure 2-2A and Table S2-2). In the primary and validation screens, Bcl-x-Venus expression correlated significantly with cell death (Figures S2-2A and S2-2C). Venus showed no systematic correlation to CFP, indicating that nonspecific promoter effects were not a major source of positives (Figures S2-2B and S2-2F). In the validation screen, apoptosis was tracked by Annexin-V-Cy5 staining and showed strong coupling to proapoptotic *Bcl-x* splicing and cell death (Figures S2-2D and S2-2E). Functional enrichments, notably splicing and cell-cycle regulation, were similar between validated hits and the primary screen (Figure 2-2C). Subsequent analyses focused on "high-confidence" factors validated with two or more siRNAs. We focused specifically on aurora kinase A (AURKA) and other mitotic regulators because of their strong functional enrichment and their novelty in the context of splicing regulation.

Physiological function for screen hits was confirmed by analysis of endogenous *Bcl-x* AS in HeLa cells. siRNAs against 19 factors were tested, and 16 (~85%) significantly shifted *Bcl-x* splicing toward *Bcl-xS* (Figure S2-3A). The magnitude of these shifts matched or exceeded fold changes that promote apoptosis in various cell types,

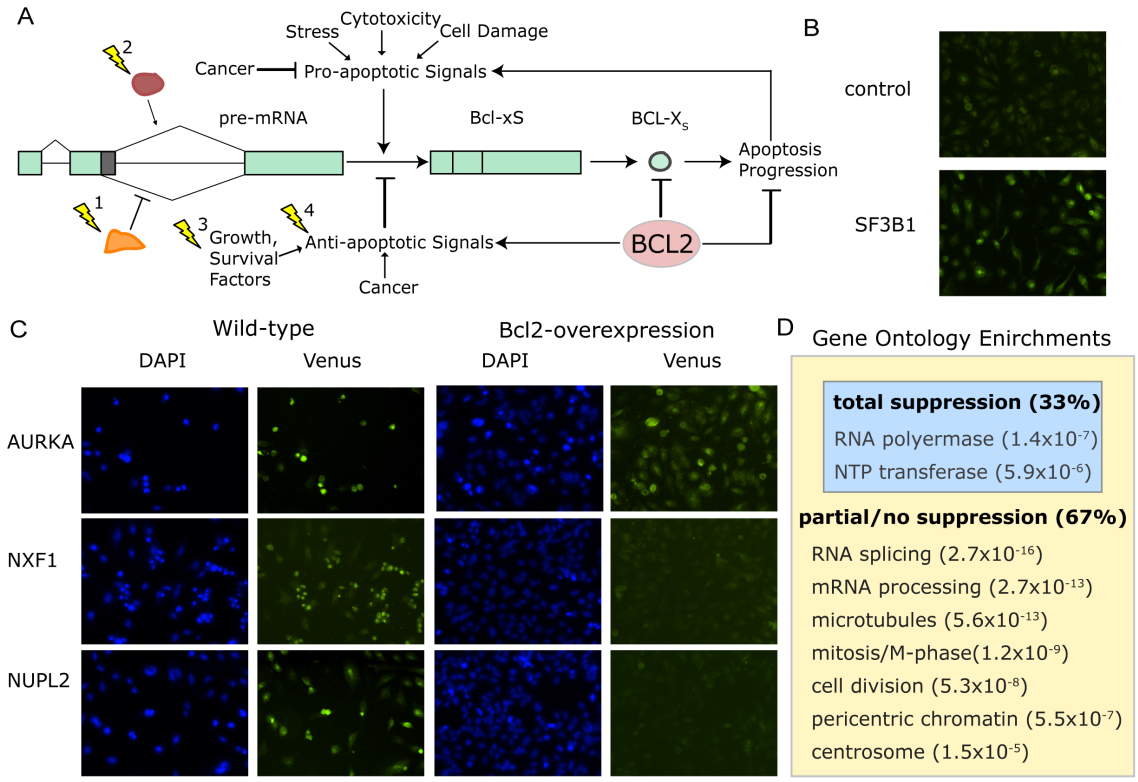

Figure 2-3: Apoptosis inhibition by Bcl2 suppressed a subset of *Bcl-x* regulators.

(A) Screen hits (lightning bolts) may affect *Bcl-x* AS by many potential mechanisms. Regulation may be direct, such as (1) direct suppression of *Bcl-xS* formation or (2) facilitation of *Bcl-xL*. Indirect or downstream regulators may (3) promote cell survival pathways or (4) suppress proapoptotic pathways. We predicted that Bcl2 overexpression would suppress the splicing regulatory effect of siRNAs targeting indirect mechanisms, but not direct regulators.

(B) Bcl2 overexpression did not suppress the effect of SF3B1 depletion, consistent with direct regulation of *Bcl-x* splicing.

(C) Bcl2 suppressed the effect of siRNAs against mRNA export factors NXF1 and NUPL2, but not AURKA, on *Bcl-x* splicing (compare Venus images between "wild-type" and "Bcl2 overexpression"). By contrast, Bcl2 suppressed cell death in all cases (compare DAPI images).

(D) Thirty-three percent of hits were suppressed by Bcl2 overexpression, while 67% showed partial to no suppression. GO functional enrichments within these subsets are shown with p values (see Table S2-3). See also Table S2-3.

indicating physiological significance (Mercatante et al. 2002; Taylor et al. 1999). As

further indication of this significance, AURKA knockdown strongly shifted endogenous Bcl-x protein toward Bcl-xS (Figure S2-3B). This shift was larger than the corresponding mRNA shift, suggesting an amplification effect during translation. Regulation of endogenous *Bcl-x* by AURKA and other hits also confirmed in MCF7 (breast adenocarcinoma) and PANC1 (pancreatic carcinoma) cells (Figure S2-3C). These lines differ markedly in steady state AURKA levels, with MCF7 (like HeLa) expressing high levels, and PANC1 levels closer to "normal" tissue (Ross et al. 2000). Thus, splicing regulation by AURKA inhibition was not restricted to HeLa cells and occurred irrespective of steady state AURKA levels.

Apoptosis suppression defines direct regulators of Bcl-x alternative splicing

In screening experiments, proapoptotic *Bcl-x* splicing correlated to apoptosis induction and increased cell death (Figure S2-2). However, screen hits are likely to include both direct regulators of *Bcl-x* splicing and factors that influence splicing indirectly via upstream activation of proapoptotic pathways (Figure 2-3A). To define systematically the relationship of screen hits to apoptosis, we retested all 274 validated regulators in *Bcl-x* reporter cells engineered to overexpress Bcl2. Bcl2 overexpression suppresses mitochondrial permeablization and hence canonical apoptosis progression (Kroemer 1997). Importantly, Bcl2 directly antagonizes the function of Bcl-xS, so this strategy also suppresses potential downstream effects of Bcl-xS-Venus produced by the reporter construct. We predicted that Bcl2 overexpression would attenuate the effects of proapoptotic siRNAs that alter *Bcl-x* splicing as part of a general proapoptotic response, but not siRNAs that target more direct regulators of splicing. Supporting the latter

prediction, depletion of direct regulator SF3B1 increased reporter expression despite Bcl2 overexpression (Figure 2-3B).

Thirty-three percent of 160 high-confidence *Bcl-x* regulators were totally suppressed by Bcl2 overexpression (Table S2-3). For example, Bcl2 suppressed both cell death and changes in Bcl-x splicing in response to depleting mRNA export factors NXF1 or NUPL2 (Figure 2-3C). By contrast, 67% of factors, including AURKA, showed partial or no suppression of proapoptotic *Bcl-x* splicing, despite the fact that cell counts showed efficient suppression of cell death. Thus, AURKA, not previously implicated in splicing regulation, functioned independently of Bcl2, whereas mRNA export factors, which directly interact with the spliceosome, were suppressed. Overall, most mRNA splicing and cell-cycle regulators showed no or partial suppression (Figure 2-3D). Suppressed factors were enriched for functions in transcription, while signaling factors spanned both categories.

Discovery of Mcl1 alternative splicing regulators

To explore functional coordination of apoptotic AS events, we determined all *Bcl-x* regulators that also control functionally analogous AS of *Mcl1*. Fifty-two of 160 "high-confidence" factors regulated both events, and we validated a subset on endogenous *Mcl1* (Figure 2-4A, Figure S2-4, and Table S2-4). Cell-cycle functions were enriched for factors that coregulated *Bcl-x* and *Mcl1*, but not factors that regulated only *Bcl-x* (Figure 2-4A). Similarly, common regulators were disproportionately enriched for splicing functions versus those that only regulated *Bcl-x*. This result indicates extensive coregulation of *Bcl-x* and *Mcl1* by cell-cycle and splicing factors. In contrast,

Figure 2-4: Coordinated regulation of *Bcl-x* and *Mcl1* alternative splicing.

(A) *Bcl-x* screen hits were tested for *Mcl1* regulation to identify common regulators. High-confidence hits are depicted as a Venn diagram, with GO enrichments listed with p values in the subsets where they were most significantly enriched (see Figure S2-4 and Table S2-4).

(B) Spliceosome-associated proteins (SAPs) that were *Bcl-x* screen hits are shown with established physical interactions and functional roles. Blue ellipses represent known RNP or protein complexes; pink ellipses contain functionally or structurally related factors. Low-confidence hits are italicized. Factors in gray rectangle have unknown roles in splicing but were identified in proteomic analyses or are similar to known splicing factors (Chen et al., 2007). Factors are marked as follows: *, high-confidence hit for Mcl1; underlined, phosphorylated in mitosis; +, putative AURKA target; x, putative CDK1 target; and #, putative PLK1 target (Dephoure et al., 2008).

(C) Hits from a kinase/phosphatase screen for *Mcl1* splicing regulation are shown, with corresponding *Bcl-x* regulators. A cutoff of $p < 10^{-5}$ in three replicates, or $p < 10^{-7}$ in two replicates was applied (see Table S2-5). Cell cycle factors are underlined. See also Figure S2-4 and Tables S2-4 and S2-5.

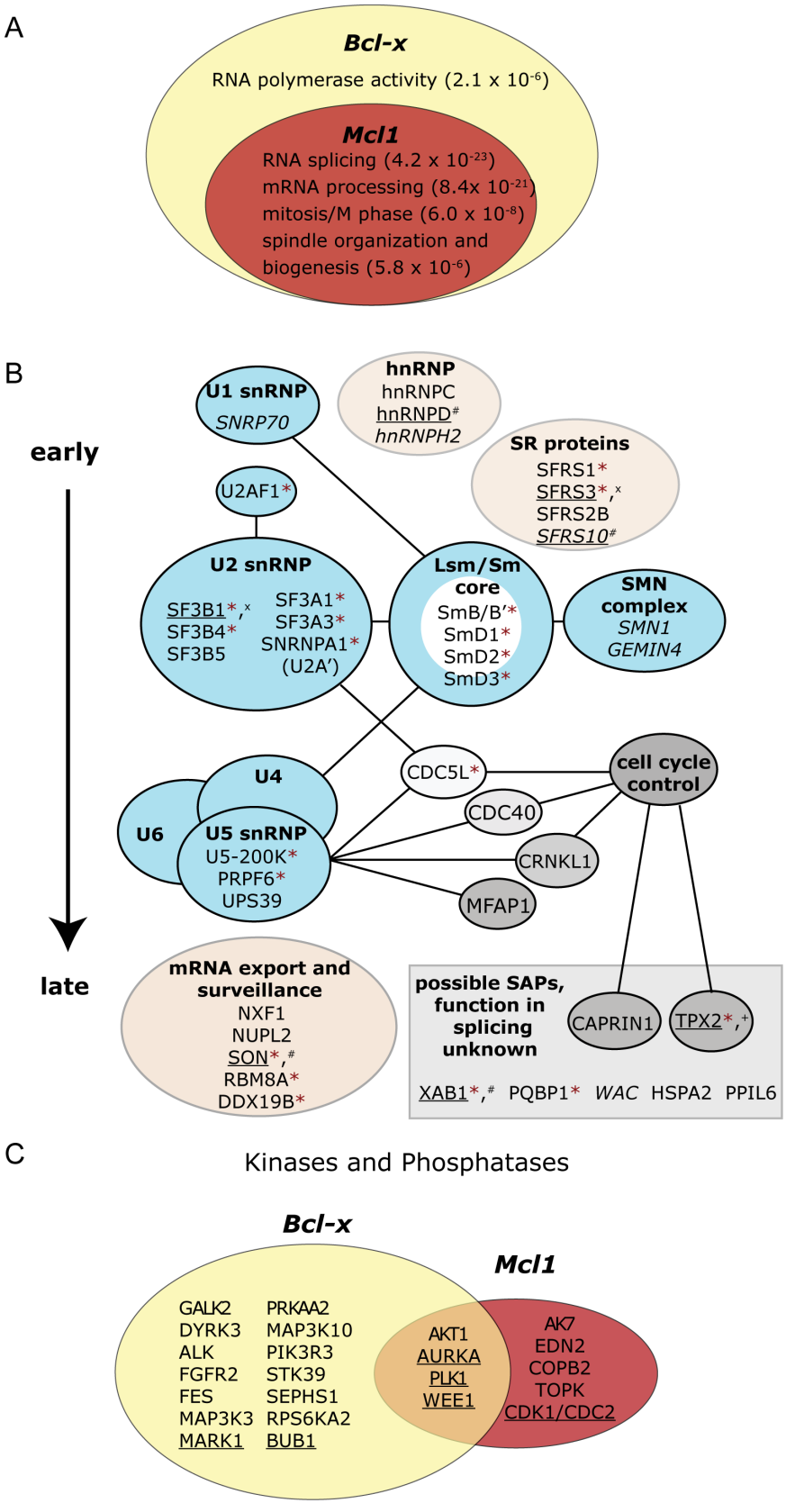

Figure 2-4 (Continued).

transcription regulators were enriched only for *Bcl-x*. Figure 2-4B schematizes hits in the *Bcl-x* and *Mcl1* screens with known or postulated functions in mRNA metabolism. Hits included "core" spliceosome components in Sm ring domains and snRNPs, as well as hnRNP and SR proteins. Interestingly, factors acting late in spliceosome assembly, including U5 snRNP components and factors previously linked to cell-cycle regulation (e.g., CDC40, CDC5L), scored positively for both *Bcl-x* and *Mcl1* (Makarova et al. 2004).

As an unbiased analysis of *Mcl1* AS, we tested siRNA pools for >700 human kinases and phosphatases in the *Mcl1* screening assay. Common regulators for *Bcl-x* and *Mcl1* included cell- cycle kinases AURKA, PLK1, and WEE1 (Figure 2-4C and Table S2-5). Further, BUB1 scored strongly for *Bcl-x* and weakly positive for *Mcl1*, though the latter did not surpass the confidence threshold. Conversely, CDK1 (CDC2) was strongly positive for *Mcl1*, and weakly so for *Bcl-x*. These data independently confirm *Mcl1* AS regulation by cell-cycle disruption.

Coupling of cell-cycle control and alternative splicing

The screen identified many cell-cycle factors, including cancer therapy targets AURKA, PLK1, and survivin (BIRC5). Protein interaction network analysis of screen hits revealed AURKA-centered interactions spanning the cell cycle, spliceosome, and tumor suppressors (Figure 2-5A). As independent validation of this regulation, nocodazole, an inhibitor of microtubule polymerization, and aurora kinase inhibitors ZM447439 and VX-680 strongly induced *Bcl-xS* formation in a dose-dependent manner (Figure 2-5B). By contrast, the broad-spectrum kinase inhibitor staurosporine induced

Figure 2-5: Alternative splicing regulation is coupled to cell-cycle control.

(A) Screen hits span an AURKA-centered protein interaction network. Circles ("nodes") depict proteins, connected by lines ("edges") representing validated protein-protein interactions. A maximum of two nonhit bridges was permitted to connect hits to AURKA, as described in the Supplemental Experimental Procedures.

(B) Drugs disrupting the cell-cycle promote proapoptotic *Bcl-x* AS. *Bcl-x* reporter cells were treated with nocodazole, aurora inhibitors VX-680 or ZM447439, or staurosporine for 18 hr. 3-fold dilution series were tested with the following maximum concentrations: 200 nM nocodazole, 10 μM VX-680, 30 μMZM447439, and 33 nM staurosporine. Percent-Venus-positive values for drug treatments were normalized to DMSO treatments; plotted values reflect means of 8–12 measurements± SD. Nocodazole, VX-680, and ZM447439, but not staurosporine, caused significant proapoptotic shifts in *Bcl-x* splicing ($p < 0.001$). LD_{50} ranges derived from cell counts are shown below the x axis. Wells with >33 nM staurosporine had >95% death, so they were not quantified (see Figure S2-5).

(C) Aurora inhibitors induced mitotic arrest coupled with proapoptotic *Bcl-x* splicing. *Bcl-x* reporter cells were treated with 3 μM VX-680, 10 μM ZM447439, or DMSO, stained with propidium iodide, and analyzed for DNA content and Venus fluorescence by flow cytometry.

(D) Asynchronous, actively cycling cells from (C) show little to no *Bcl-x* splicing fluctuations relative to arrested cells.

(E) S phase arrest by double thymidine block did not alter *Bcl-x* splicing, indicating specificity for mitotic arrest. Cells were analyzed as in (C).

(F) *Bcl-x* splicing regulation precedes mitotic arrest. Asynchronous or thymidine-synchronized cells from (E) were treated with the indicated inhibitors or DMSO. Cells were analyzed by automated microscopy 16 hr after thymidine release, precluding an intervening round of mitosis for synchronized cells. The response in presynchronized cells confirms a splicing shift upstream of mitotic arrest. See also Figure S2-5.

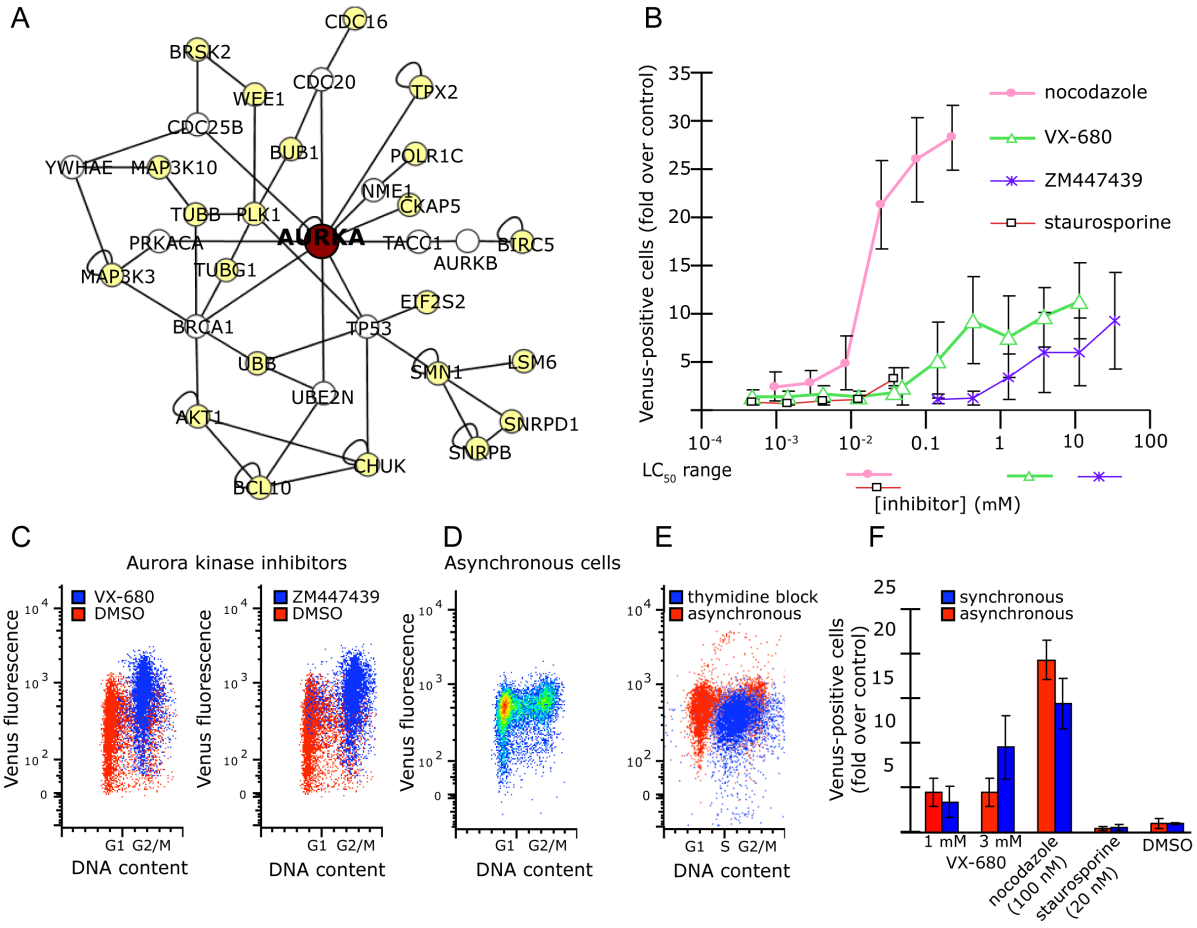

Figure 2-5 (Continued).

apoptosis but not *Bcl-xS* formation, demonstrating specific regulation by cell-cycle inhibitors. ZM447439 and VX-680 inhibit the three mammalian aurora kinases AURKA, AURKB, and germline-restricted AURKC. However, ZM447439 is ~20-fold more potent toward AURKB than AURKA in vitro, while VX-680 is ~30-fold more potent against AURKA (Ditchfield et al. 2005; Harrington et al. 2004). Since *Bcl-x* splicing regulation in screening assays was AURKA specific, VX-680 was used subsequently because of its higher specificity for AURKA. Importantly, nocodazole and VX-680 treatment also shifted endogenous *Bcl-x* and *Mcl1* splicing (Figure S2-5).

In flow cytometry experiments, aurora inhibitors induced mitotic arrest concomitant with upregulation of the *Bcl-x* reporter (Figure 2-5C). Bcl-xS-Venus levels in actively cycling G2/M cells were slightly higher than in G1, but the difference was not significant compared to arrested cells (Figure 2-5D). In addition, S phase arrest by double-thymidine block did not affect Bcl-xS-Venus levels, so the splicing response was specific to mitotic arrest (Figure 2-5E). Since transcription and splicing are primarily silenced in M phase (Shin and Manley 2002), we considered two possibilities: (1) splicing changes preceded arrest or (2) a subset of cells underwent an abnormal division prior to full arrest, leading to a proapoptotic stress response. To test the latter scenario, cells were synchronized by double-thymidine block, released, and treated with VX-680. Venus fluorescence was measured 16 hr after thymidine release, thus ruling out an intervening mitotic cycle. Figure 2-5F shows the splicing response to VX-680 in synchronized cells was comparable to asynchronous cells. Apoptosis induction by staurosporine caused no splicing response, regardless of synchronization. These results show that stimuli that activate G2/M and spindle checkpoints trigger proapoptotic *Bcl-x*

splicing, but the response precedes actual arrest.

Downstream mechanisms of splicing regulation

To identify direct splicing regulators downstream of cell-cycle inhibition, we examined expression of splicing factors upon AURKA knockdown, focusing on strong screen hits or factors previously linked to *Bcl-x* splicing (Figure S2-6A). Among these factors, only ASF/SF2 (SFRS1)—an SR protein previously linked to cell-cycle regulation and a strong hit in the primary screen— showed specific downregulation upon AURKA knockdown (Figure 2-6A) (Li et al. 2005). Inhibition of AURKA with VX-680 also resulted in dose-responsive downregulation of ASF/SF2 (Figure 2-6B).

ASF/SF2 mRNA was not significantly reduced upon AURKA inhibition, ruling out transcriptional downregulation or effects on RNA stability (Figure S2-6B). However, VX-680 treatment reduced exogenous green fluorescent protein (GFP)-tagged ASF/SF2 in a manner similar to endogenous ASF/SF2, indicating posttranslational turnover as the main source of ASF/SF2 depletion (Figure 2-6C). AURKA inhibition also induced ASF/SF2 downregulation in apoptosis-resistant cells that overexpress Bcl2 (Figure S2-6C). Therefore, this regulatory pathway functions in cells where apoptosis is suppressed, consistent with our finding that AURKA knockdown shifts *Bcl-x* splicing when apoptosis is suppressed (Figure 2-3C).

To determine whether *Bcl-x* is a direct target of ASF/SF2, we used the CLIP method (ultraviolet [UV] crosslinking and immunoprecipitation) to analyze RNAs bound directly by ASF/SF2 in vivo. RNP complexes were covalently crosslinked in living cells by UV exposure, and ASF/SF2-RNA complexes were immunopurified (Figure 2-6D).

Figure 2-6: ASF/SF2 regulates splicing downstream of AURKA.

(A) siRNA depletion of AURKA downregulated ASF/SF2. Cells were transfected with two different siRNAs for AURKA or AURKB, or control siRNAs. Lysates were analyzed by western analysis for indicated factors. No other splicing factors measured in these samples showed significant changes (Figure S2-6A), indicating specificity for ASF/SF2.

(B) VX-680 treatment downregulated ASF/SF2 in a dose-dependent manner.

(C) ASF/SF2 is downregulated posttranslationally upon AURKA inhibition (see also Figure S2-6B). HeLa cells were transfected with a GFP-ASF/SF2 construct, and then treated with VX-680 or DMSO. Western analysis showed that VX-680 downregulated exogenous ASF/SF2, indicating posttranslational turnover. ASF/SF2 downregulation also occurred in apoptosis-resistant cells overexpressing Bcl2 (see Figure S2-6C).

(D) Visualization of ASF/SF2-RNA complexes by CLIP is shown. RNP complexes were UV crosslinked in live HeLa cells, ASF/SF2 was immunopurified, and RNA was end labeled with γ-^{32}P-ATP. Complexes were run on SDS-PAGE and visualized by autoradiography. Noncrosslinked ASF/SF2 (UV– lane) was also labeled, as observed in Sanford et al. (2009).

(E) ASF/SF2 directly binds *Bcl-x* and *Mcl1* mRNAs. RT-PCR analysis of ASF/SF2-bound RNAs showed significant enrichment of *Bcl-x*, *Mcl1*, and known target *TPX2* versus an irrelevant IgG control. Values are the means of three independent measurements ± SD.

(F) Exogenous expression of ASF/SF2 attenuated the effect of AURKA inhibition on *Bcl-x* splicing. *Bcl-x* reporter cells were transfected with mCherry-tagged ASF/SF2, a mutant lacking the second RRM domain (ΔRRM2), or mCherry alone. Eighteen hours after transfection, cells were treated with 3 μm VX-680 or DMSO for 24 hr. The fraction of Venus-positive cells among mCherry-positive cells was determined by automated microscopy. VX-680-treated samples were background corrected by subtraction of values from DMSO treatment. Values are the means of three separate measurements ± SD, expressed relative to the mCherry-transfected control. ASF/SF2-WT attenuated the *Bcl-x* response to VX-680 by ~50% versus ΔRRM2. CFP expression was not affected, indicating specific attenuation of *Bcl-x* splicing.

(G) ASF/SF2-ΔRRM2 shows deficient association with *Bcl-x* and *Mcl1* mRNAs. CLIP analysis of exogenous, GFP-tagged ASF/SF2-WT or -ΔRRM2 was performed with α-GFP antisera. Values are the means of three independent measurements ± SD. See also Figure S2-6.

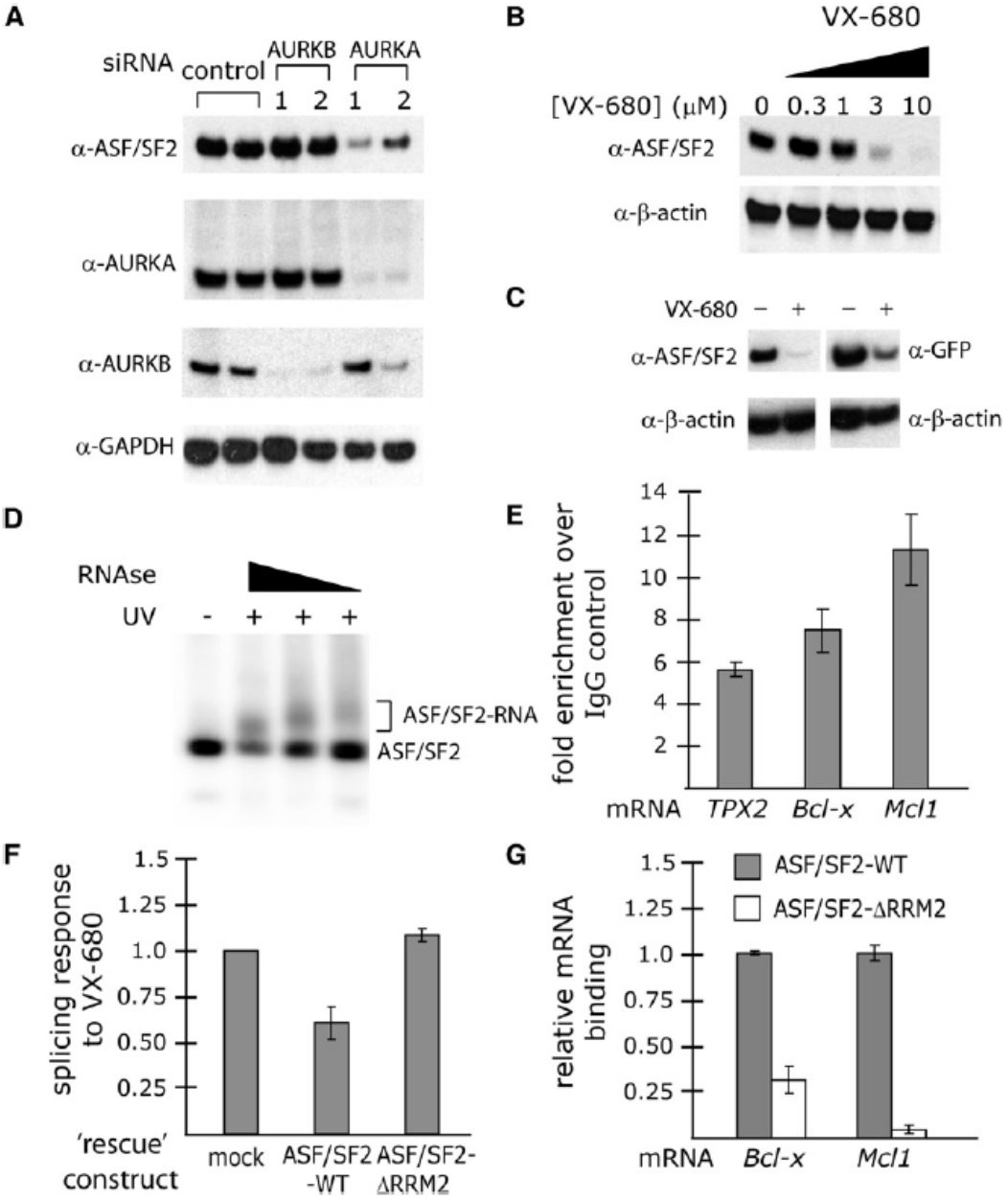

Figure 2-6 (Continued).

RT-PCR analysis showed significant enrichment of *Bcl-x* mRNA in ASF/SF2-RNA immunoprecipitations (IPs) versus an irrelevant IgG control, along with *Mcl1* and previously known target *TPX2* (Figure 2-6E) (Sanford et al. 2009). *APAF1* and *TNFSF13* mRNAs were detectable in input but not ASF/SF2 IPs, demonstrating that crosslinking was specific.

Our data show direct regulation of *Bcl-x* and *Mcl1* splicing by ASF/SF2 and downregulation of ASF/SF2 by AURKA inhibition. To test whether ASF/SF2 functions downstream of AURKA inhibition, we examined whether exogenous ASF/SF2 expression could "rescue" the splicing response to AURKA inhibition. *Bcl-x* reporter cells were transfected with mCherry-tagged constructs expressing wild-type (WT) ASF/SF2 or a variant lacking the second RNA recognition motif (ΔRRM2) domain. After VX-680 treatment, the *Bcl-x* splicing response in cells expressing ASF/SF2-WT was reduced ~50% relative to cells expressing ΔRRM2 or mCherry alone (Figure 2-6F). CLIP of exogenous GFP-tagged ASF/SF2 constructs confirmed deficient recognition of endogenous *Bcl-x* mRNA by the ΔRRM2 mutant (Figure 2-6G) (Cáceres et al. 1997). Therefore, enforced ASF/SF2 expression rescued the effect of AURKA inhibition on *Bcl-x* splicing, and rescue was dependent on intact splicing regulatory function.

ASF/SF2 regulates apoptosis globally in response to cell-cycle inhibition. Consistent with previous reports, siRNA depletion of ASF/SF2 activates markers of apoptosis (e.g., Annexin-V staining, DNA fragmentation, and caspase cleavage) (data not shown) (Li et al. 2005; Karni et al. 2007). To analyze apoptosis regulation by ASF/SF2 in the context of cell-cycle disruption, we depleted ASF/SF2 by RNAi and analyzed apoptosis markers upon AURKA inhibition with VX-680. ASF/SF2 depletion enhanced

caspase activation and Annexin-V staining relative to a nontargeting siRNA control (Figures 2-7A and 2-7B). In addition, we compared the effects of ASF/SF2 depletion to the depletion of two splicing factors (SRPK1 and SNRP70) that had no significant effect on endogenous *Bcl-x* splicing and a third factor (Sam68) that has antagonistic effects to ASF/SF2 on *Bcl-x* splicing (Figure S2-3A and Table S2-1) (Paronetto et al. 2007). ASF/SF2 depletion sensitized cells to VX-680-induced apoptosis relative to SRPK1, SNRP70, or Sam68 depletion. In contrast, ASF/SF2 depletion did not enhance staurosporine-induced apoptosis relative to SRPK1, SNRP70, or Sam68 depletion (Figure S2-6D). We could not reliably analyze the effects of enforced ASF/SF2 expression on apoptotic markers because overexpression was toxic in the time frame required for this analysis. In sum, ASF/SF2 loss-of-function triggers apoptosis and specifically sensitizes cells to apoptosis induced by AURKA inhibition.

To examine broader functions of AS in this apoptotic response, we analyzed *caspase-9* (CASP9) and *caspase-2* (CASP2) mRNAs, which produce short (S), antiapoptotic, or long (L), proapoptotic isoforms (Schwerk and Schulze-Osthoff 2005). Depletion of ASF/SF2 and AURKA favored proapoptotic *CASP9* splicing (Figure 2-7C) (Massiello and Chalfant 2006). In addition, ASF/SF2 depletion favored proapoptotic *CASP2* splicing, although AURKA knockdown had no significant effect. We extended our analysis to additional factors with diverse apoptotic functions and AS patterns, including exon skipping, alternative 5' ss use, and intron retention. Splice isoforms were measured with an isoform-sensitive qPCR strategy upon knockdown of various screen hits, revealing widespread changes in splicing upon silencing of cell-cycle regulators (Figures 2-7D). This approach is only semiquantitative, so we focused on identifying

Figure 2-7: An apoptotic alternative splicing program linked to cell-cycle control.

(A) ASF/SF2 downregulation triggers apoptosis and sensitizes cells to apoptosis induction by AURKA inhibition. HeLa cells were transfected with indicated siRNAs, incubated to allow factor depletion, and then treated with varying concentrations of VX-680. Apoptosis was measured via cleavage of a fluorescent caspase-3 substrate. Depletion of ASF/SF2, but not SRPK1, SNRP70, or Sam68, sensitized cells to VX-680-induced apoptosis. This selective sensitization by ASF/SF2 depletion was not observed for staurosporine-induced apoptosis (see Figure S2-6D). Data reflect the means of three independent transfections± SD.

(B) Analysis in (A) was repeated with Annexin-V staining used as the apoptosis marker.

(C) *CASP9* and *CASP2* AS were analyzed in HeLa cells by qPCR after transfection with the indicated siRNAs. Relative levels of proapoptotic isoform (L) normalized to negative controls are shown. Values are the means of three independent measurements ± SD.

(D) Results of AS profiling by real-time qPCR are shown in a heat-map of Z scores averaged across four biological replicates, along with unsupervised hierarchical clustering (see Figure S2-7 for an expanded view and validation of these data).

(E) A model of cell-cycle arrest coupled to apoptosis via AS is shown. We propose anticipatory, proapoptotic splicing decisions in interphase that later promote apoptosis in arrested cells. Lightning bolts represent siRNAs, drugs, or other stimuli that disrupt the cell cycle and promote proapoptotic splicing of *Bcl-x*, *Mcl1*, *CASP9*, and other targets. See also Figures S2-6 and S2-7.

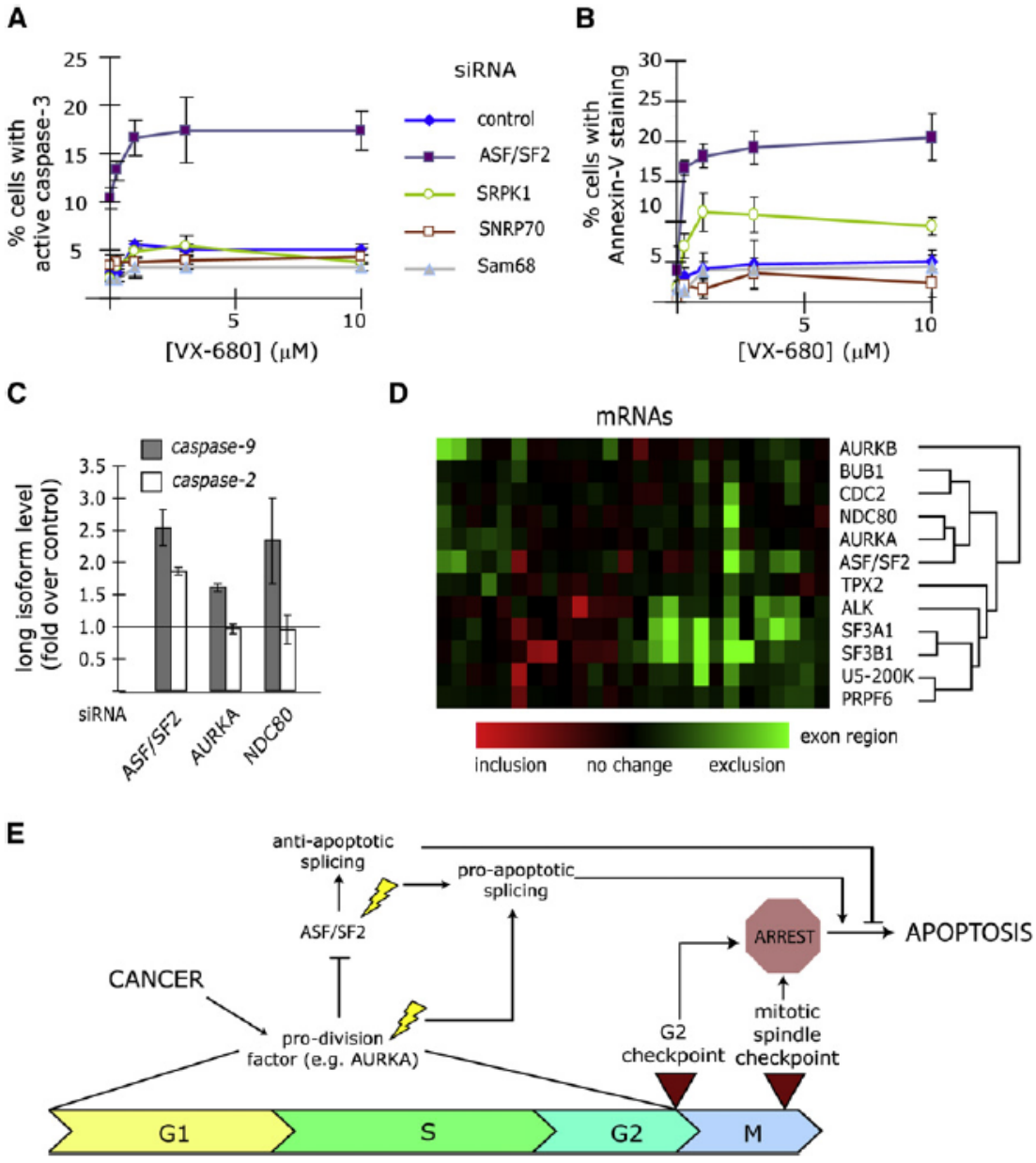

Figure 2-7 (Continued).

biologically robust events by Z score analysis of four independent experiments (Figures S2-7A and S2-7B). Hierarchical clustering of knockdown profiles revealed overlapping regulation by cell-cycle kinases AURKA, CDC2, and BUB1. In addition, splicing factor profiles grouped by validated spliceosomal interactions; the U2 snRNP factors SF3B1 and SF3A1 overlapped significantly, as did U5 factors PRPF6 and U5-200K. Interestingly, ASF/SF2 targets clustered with AURKA and kinetochore component NDC80, suggesting overlapping regulation in this specific set of mRNAs. As confirmation of direct regulation, binding of ASF/SF2 to selected targets *CASP9*, *CASP2*, *ATG4*, and *BMF* was established by CLIP (Figure S2-7D).

DISCUSSION

Genome-wide identification of alternative splicing regulators in human cells

We developed a strategy for systematic, high-throughput analysis of AS regulation in human cells. Our siRNA screens identified many spliceosomal regulators of *Bcl-x* and *Mcl1* splicing, as well as functional connections to cell-cycle control, cytoskeleton dynamics, and signal transduction (Figure 2-2C). Strong functional enrichment for known splicing factors was an encouraging indication of specificity. In addition, a high validation rate of screen hits on endogenous AS indicated faithful modeling of physiological regulation in our reporter assays (Figure S2-3). Broadly, our efforts expand the catalog of factors linking mRNA metabolism and apoptosis and provide a platform for future high-throughput analyses of mRNA processing in living cells.

Screen hits spanned essential processes, such as transcription, cell division, and

prosurvival signaling, with extensive mechanistic links to apoptosis, implicating splicing regulation in diverse proapoptotic pathways. The strong correlation of *Bcl-xS* formation with apoptosis induction and cell death further support a physiological function for splicing regulation in apoptosis (Figure S2-2). To investigate the functional connection between splicing regulation and apoptosis, we retested factors in cells rendered resistant to apoptosis by exogenous Bcl2 expression (Figure 2-3). Surprisingly, cell-cycle factors, like most spliceosomal factors, were generally not affected by Bcl2 expression. Therefore, functional coupling of cell-cycle disruption to proapoptotic splicing remained intact in cells in which the canonical apoptosis pathway was suppressed. This finding suggests that coordinated, proapoptotic splicing could be exploited to target apoptosis-resistant cell populations, such as tumors.

Our screening experiments also revealed functional coordination of proapoptotic splicing events. Fifty-two of 160 "high-confidence" *Bcl-x* regulators affected functionally analogous splicing of *Mcl1*, and common regulators were disproportionately enriched for splicing and cell-cycle functions (Figure 2-4A). Overlapping spliceosomal regulators of *Bcl-x* and *Mcl1* included "alternative splicing factors" (e.g., hnRNPs and SR proteins) and "core" spliceosome components spanning most steps in spliceosome assembly (Figure 2-4B). Interestingly, U5 snRNP and associated factors that function late in spliceosome assembly affected splice-site choice for *Bcl-x*, *Mcl1*, and other mRNAs (Figures 2-4B, 2-7C, and 2-7D). Current mechanistic models support splice-site pairing early in spliceosome assembly upon U2 snRNP recruitment (Dassah et al. 2009). Our findings suggest that perturbations of late steps in spliceosome function may promote remodeling of splice-site choice downstream of the rate-limiting steps that normally

determine splicing decisions.

Cell-cycle disruption coupled to proapoptotic alternative splicing

Our screen for AS regulators identified factors involved in mitotic spindle assembly, kinetochore and centromere dynamics, microtubule transport, centrosome duplication and dynamics, and cell division (Figure 2-5A). Well-characterized pharmacological inhibitors, notably for AURKA, confirmed functional coupling of cell-cycle and splicing regulation (Figure 2-5B and Figure S2-5). In flow cytometry experiments, proapoptotic *Bcl-x* splicing was coupled to mitotic arrest, but not S phase arrest or G2/M in actively cycling cells (Figures 2-5C–2-5E). Since prevailing evidence supports suppression of transcription and splicing in mitosis, it is unlikely that the observed response initiates in mitotically arrested cells (Shin and Manley 2002). Moreover, when cells were presynchronized to preclude an intervening cell division, VX-680 still induced a robust splicing response (Figure 2-5F). Therefore, this splicing regulation precedes mitotic arrest, indicating accumulation of prodeath factors in anticipation of G2/M and/or mitotic spindle checkpoints. Since cell-cycle disruption also promoted a proapoptotic switch in *CASP9* splicing and altered splicing of other apoptotic regulators, this regulation likely extends to many coregulated mRNAs (Figures 2-7C and 2-7D).

Many previous studies have linked cell-cycle and splicing regulation, reaching back to the seminal isolation of cell-division cycle (CDC) loci in *S. cerevisiae* that encode spliceosome components, including factors (CDC5L and CDC40) identified here. Recent genetic and proteomic analyses of mitotic spindle assembly have reasserted the

notion that the cell-cycle and splicing machineries are intertwined, but functional consequences have not been established (Björklund et al. 2006, Rines et al. 2008). We observe coordinated proapoptotic splicing upon inhibition of key cell-cycle factors, but preceding mitotic arrest, indicating anticipatory upregulation of prodeath factors that may subsequently propagate apoptosis in arrested cells. Since arrested cells cannot mount de novo transcriptional and splicing responses, this finding suggests a genetic mechanism for cell-cycle checkpoint response.

ASF/SF2 regulates an apoptotic splicing network

We present evidence that ASF/SF2 downregulation mediates a proapoptotic splicing response to cell-cycle disruption. ASF/SF2 was a hit in the *Bcl-x* and *Mcl1* screens, and its overexpression was previously shown to favor *Bcl-xL* formation (Paronetto et al. 2007). AURKA inhibition triggered posttranslational turnover of ASF/SF2, but not other SR proteins or splicing factors identified in the screen (Figures 2-6A–2-6C and Figure S2-6A). In addition, ASF/SF2 directly binds *Bcl-x* and *Mcl1* mRNAs, and its enforced expression attenuated the effect of AURKA inhibition on *Bcl-x* splicing (Figures 2-6D–2-6G).

ASF/SF2 regulation extends beyond *Bcl-x* to global regulation of apoptosis (Li et al. 2005). ASF/SF2 depletion, but not depletion of several splicing factors that were not validated screen hits, sensitized cells to apoptosis induction by AURKA inhibition (Figures 2-7A and 2-7B). In contrast, ASF/SF2 depletion did not specifically enhance apoptosis in response to staurosporine (Figure S2-6D). ASF/SF2-mediated splicing regulation extended to other functional targets, including *CASP9*, which underwent a

proapoptotic splicing shift upon depletion of AURKA and ASF/SF2 (Figure 2-7C). Our analysis of some 25 additional AS targets revealed overlapping effects of AURKA and ASF/SF2 inhibition (Figure 2-7D). The fact that some mRNAs (e.g., *CASP2*) did not respond identically to AURKA and ASF/SF2 knockdown may reflect differential effects of targeting upstream versus direct regulators of splicing, or the activity of additional, unidentified splicing regulators. Nonetheless, these experiments strongly indicate that ASF/SF2 downregulation globally shifts alternative splicing patterns to promote apoptosis upon disruption of the cell cycle.

Alternative splicing in a global context

Our data link coordinated proapoptotic splicing to activated cellcycle checkpoints (Figure 2-7E). These findings are relevant to basic mechanisms of apoptosis in arrested cells, which are poorly defined, and the exploitation of these pathways in cancer therapy. Aberrant expression of AURKA and other prodivision factors promotes spindle checkpoint abrogation, chromosomal instability, and aneuploidy in tumors (Lapenna and Giordano 2009). Our findings indicate that AURKA activation may also propagate oncogenesis by promoting antiapoptotic splicing of crucial cell-death regulators. Conversely, AURKA inhibition elicits parallel upregulation of prodeath variants, which may contribute to its demonstrated efficacy as a cancer drug target. Importantly, this regulation remains intact in apoptosis-resistant cells, suggesting functional and potential therapeutic relevance in tumors.

ASF/SF2 downregulation triggers proapoptotic splicing and promotes apoptosis. Importantly, ASF/SF2 is a proto-oncogene in its own right, and its depletion is linked to

genome instability and G2/M arrest (Karni et al. 2007; Li and Manley 2005). Our data highlight RNA targets that may contribute to its transforming properties. Taken together, these findings suggest that ASF/SF2 has multiple, synergistic functions in controlling cell arrest and apoptosis, underscoring its role as a central regulator of mRNA metabolism in cancer. Finally, it bears reiterating that our systematic screening approach was the original source of these biological insights, as well as many new regulators and functional connections not expounded here. We believe future screening efforts will be instrumental in defining global roles for AS in physiological and disease contexts.

MATERIALS AND METHODS

siRNA screening

Cells were transfected in triplicate in 384-well plates with >21,000 siRNA SMARTpools (Thermo-Fisher) for known and predicted human genes. Seventy-two hours later, cells were fixed, DAPI stained, and imaged on the Image Xpress Micro microscope (Molecular Devices). Cell counts (DAPI), percent of Venus-positive cells, and average per-cell CFP signal were determined with MetaXpress software. Wells were then assigned probabilistic "strength" and "confidence" scores in the Venus channel using a support vector machine (SVM) model trained on positive and negative control data (Figure S2-1, Table S2-1, and the Supplemental Materials). For independent verification, plate-wide, median-normalized Z scores were also tabulated. Wells with SVM strength scores with false discovery rate (FDR) <10, or $z > 2.6$ were inspected visually.

Three hundred and sixty-nine positives were retested in a validation screen with

four siRNAs from deconvoluted SMARTpools. siRNAs >2 standard deviations (SDs) above negative controls in the Venus channel were deemed positive (Table S2-2). Bcl2-overexpression and *Mcl1* screens were performed and analyzed as for the validation screen (Tables S2-3 and S2-4).

RNA extraction and RT-PCR

Cellular RNA was extracted with Trizol (Invitrogen) reagent, and 2.5–5 μg were reverse transcribed with Superscript III (Invitrogen) and oligo-dT. *Bcl-x*, *Mcl1*, *CASP9*, and *CASP2* isoforms were analyzed by qPCR with primers flanking alternative exon sequences. For each product, a linear range of cDNA input was determined empirically by running of a 2-fold dilution series in 26–28 cycle reactions. Products were run on 5% TBE-PAGE gels, stained with Sybr-Gold (Invitrogen), and quantified with QuantityOne software (Biorad).

Real-time qPCR primer sets were designed with Primer3 webware and tested by electrophoresis to verify correctly sized products (Table S2-6). Reactions were prepared with Sybr-Green PCR mix (ABI), run on the ABI 7900HT system, and quantified by a ΔΔC(t) method. Amplicons from alternative regions were first normalized to a constitutive amplicon from the same transcript. To correct for differential primer efficiencies, data were then normalized to negative controls. Z scores were tabulated across four independent experiments and clustered with Eisen Cluster 2.0.

Splicing assays with drug treatments

Sources for inhibitors are in the Supplemental Materials. Cells were treated in 96-

well plates for 18–24 hr and analyzed by automated microscopy. For flow cytometry, cells were fixed in 3% PFA/1X PBS then 70% ethanol, permeablized in 1X PBS/0.1% Triton X-100, and stained with 20 μg/ml propidium iodide. DNA content and Venus fluorescence were analyzed on the LSRII system (BD Biosciences).

Western blotting

Proteins were extracted in RIPA buffer (IX PBS, 1% Triton X-100, 0.5% sodium deoxycholate, 0.1% SDS, protease inhibitors) and quantified by Protein Dye-Binding Assay (Biorad). Ten to fifty micrograms of total protein were run on 4%–12% Bis-Tris polyacrylamide gels (Invitrogen), transferred to PVDF membranes, and probed with antibodies listed in Supplemental Materials.

CLIP

UV crosslinking and RNP purification/visualization were performed as described (Sanford et al., 2009). For quantification, CLIP and input samples were treated with Proteinase K (Ambion), and RNA was extracted, reverse transcribed, and analyzed by real-time qPCR.

Apoptosis assays

Annexin-V-Cy5 (Abcam) staining was quantified by automated microscopy. Caspase activation was measured with the DEVD-Nucview 488 reagent (Biotum), which stains nuclei specifically after it is cleaved by caspase-3. Cells were incubated for 1 hr with 2.5 μM substrate, fixed, and analyzed by automated microscopy.

ACKNOWLEDGMENTS

We thank Junying Yuan and Robin Reed for reagents, Jodene Moore for help with flow cytometry, and Jessica Hurt, Natalie Gilks-Farny, Daniel Ducat, William Senapedis, Faisal Aldaye, Deborah Flusberg, and Ian Swinburne for thoughtful critiques and discussions. RNAi screening capability was provided by the ICCB-Longwood screening facility at Harvard Medical School. We are indebted to ICCB-L staff Caroline Shamu, Sean Johnston, Stewart Rudnicki, Zac Cooper, and David Wrobel. Work was supported by funding from the National Science Foundation and grant NIHGM057476 from the National Institutes of Health.

REFERENCES

Akgul C, Moulding DA, Edwards SW. 2004. Alternative splicing of Bcl-2-related genes: functional consequences and potential therapeutic applications. *Cell Mol Life Sci* **17:** 2189–2199.

Björklund M, Taipale M, Varjosalo M, Saharinen J, Lahdenperä J, Taipale J. 2006. Identification of pathways regulating cell size and cell-cycle progression by RNAi. *Nature* **439:** 1009–1013.

Black DL. 2003. Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* **72:** 291–336.

Blencowe BJ. 2006. Alternative splicing: new insights from global analyses. *Cell* **14:** 37–47.

Cáceres JF, Misteli T, Screaton GR, Spector DL, Krainer AR. 1997. Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity. *J Cell Biol* **138:** 225–238.

Chen YI, Moore RE, Ge HY, Young MK, Lee TD, Stevens SW. 2007. Proteomic analysis of in vivo-assembled pre-mRNA splicing complexes expands the catalog of participating factors. *Nucleic Acids Res* **35:** 3928–3944.

Cloutier P, Toutant J, Shkreta L, Goekjian S, Revil T, Chabot B. 2008. Antagonistic effects of the SRp30c protein and cryptic 50 splice sites on the alternative splicing of the apoptotic regulator Bcl-x. *J Biol Chem* **283:** 21315–21324.

Dassah M, Patzek S, Hunt VM, Medina PE, Zahler AM. 2009. A genetic screen for suppressors of a mutated 50 splice site identifies factors associated with later steps of spliceosome assembly. *Genetics* **182:** 725–734.

Dephoure N, Zhou C, Villén J, Beausoleil SA, Bakalarski CE, Elledge SJ, Gygi SP. 2008. A quantitative atlas of mitotic phosphorylation. *Proc Natl Acad Sci USA* **105:** 10762–10767.

Ditchfield C, Keen N, Taylor SS. 2005. The Ipl1/Aurora kinase family: methods of

inhibition and functional analysis in mammalian cells. *Methods Mol Biol* **296:** 371–381.

Fesik SW. 2005. Promoting apoptosis as a strategy for cancer drug discovery. *Nat Rev Cancer* **5:** 876–885.

Garneau D, Revil T, Fisette JF, Chabot B. 2005. Heterogeneous nuclear ribonucleoprotein F/H proteins modulate the alternative splicing of the apoptotic mediator Bcl-x. *J Biol Chem* **280:** 22641–22650.

Hardwick JM, Youle RJ. 2009. SnapShot: BCL-2 proteins. *Cell* **138:** 404, 404.e1.

Harrington EA, Bebbington D, Moore J, Rasmussen RK, Ajose-Adeogun AO, Nakayama T, Graham JA, Demur C, Hercend T, Diu-Hercend A, et al. 2004. VX-680, a potent and selective small-molecule inhibitor of the Aurora kinases, suppresses tumor growth in vivo. *Nat Med* **10:** 262–267.

Karni R, de Stanchina E, Lowe SW, Sinha R, Mu D, Krainer AR. 2007. The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nat Struct Mol Biol* **14:** 185–193.

Keene JD. 2007. RNA regulons: coordination of post-transcriptional events. *Nat Rev Genet* **8:** 533–543.

Kromer G. 1997. The proto-oncogene Bcl-2 and its role in regulation apoptosis. *Nat Med* **3:** 614–620.

Lapenna S, Giordano A. 2009. Cell cycle kinases as therapeutic targets for cancer. *Nat Rev Drug Discov* **8:** 547–566.

Letai AG. 2008. Diagnosing and exploiting cancer's addiction to blocks in apoptosis. *Nat Rev Cancer* **8:** 121–132.

Li X, Manley JL. 2005. Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell* **122:** 365–378.

Li X, Wang J, Manley JL. 2005. Loss of splicing factor ASF/SF2 induces G2 cell cycle arrest and apoptosis, but inhibits internucleosomal DNA fragmentation. *Genes Dev* **19:** 2705–2714.

Makarova OV, Makarov EM, Urlaub H, Will CL, Gentzel M, Wilm M, Lührmann R. 2004. A subset of human 35S U5 proteins, including Prp19, function prior to catalytic step 1 of splicing. *EMBO J* **23:** 2381–2391.

Massiello A, Chalfant CE. 2006. SRp30a (ASF/SF2) regulates the alternative splicing of caspase-9 pre-mRNA and is required for ceramide-responsiveness. *J Lipid Res* **47:** 892–897.

Massiello A, Roesser JR, Chalfant CE. 2006. SAP155 Binds to ceramide-responsive RNA cis-element 1 and regulates the alternative 50 splice site selection of Bcl-x pre-mRNA. *FASEB J*. **20:** 1680–1682.

Mercatante DR, Mohler JL, Kole R. 2002. Cellular response to an antisense-mediated shift of Bcl-x pre-mRNA splicing and antineoplastic agents. *J Biol Chem* **277:** 49374–49382.

Moore MJ, Silver PA. 2008. Global analysis of mRNA splicing. *RNA* **14:** 197–203.

Orengo JP, Bundman D, Cooper TA. 2006. A bichromatic fluorescent reporter for cell-based screens of alternative splicing. *Nucleic Acids Res* **34:** e148.

Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* **40:** 1413–1415.

Paronetto MP, Achsel T, Massiello A, Chalfant CE, Sette C. 2007. The RNA-binding protein Sam68 modulates the alternative splicing of Bcl-x. *J Cell Biol* **176:** 929–939.

Rines DR, Gomez-Ferreria MA, Zhou Y, DeJesus P, Grob S, Batalov S, Labow M, Huesken D, Mickanin C, Hall J, et al. 2008. Whole genome functional analysis identifies novel components required for mitotic spindle integrity in human cells. *Genome Biol* **9:** R44.

Ross DT, Scherf U, Eisen MB, Perou CM, Rees C, Spellman P, Iyer V, Jeffrey SS, Van de Rijn M, Waltham M, et al. 2000. Systematic variation in gene expression patterns in human cancer cell lines. *Nat Genet* **3:** 227–235.

Sanford JR, Wang X, Mort M, Vanduyn N, Cooper DN, Mooney SD, Edenberg HJ, Liu Y. 2009. Splicing factor SFRS1 recognizes a functionally diverse landscape of RNA transcripts. *Genome Res* **19:** 381–394.

Schwerk C, Schulze-Osthoff K. 2005. Regulation of apoptosis by alternative pre-mRNA splicing. *Mol Cell* **19:** 1–13.

Shin C, Manley JL. 2002. The SR protein SRp38 represses splicing in M phase cells. *Cell* **111:** 407–417.

Stoilov P, Lin CH, Damoiseaux R, Nikolic J, Black DL. 2008. A high-throughput screening strategy identifies cardiotonic steroids as alternative splicing modulators. *Proc Natl Acad Sci USA* **105:** 11218–11223.

Taylor JK, Zhang QQ, Wyatt JR, Dean NM. 1999. Induction of endogenous Bcl-xS through the control of Bcl-x pre-mRNA splicing by antisense oligonucleotides. *Nat Biotechnol* **11:** 1097–1100.

Wahl MC, Will CL, Lührmann R. 2009. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136:** 701–718.

Wang C, Youle RJ. 2009. The role of mitochondria in apoptosis. *Annu Rev Genet* **43:** 95–118.

Warzecha CC, Sato TK, Nabet B, Hogenesch JB, Carstens RP. 2009. ESRP1 and ESRP2 are epithelial cell-type-specific regulators of FGFR2 splicing. *Mol Cell* **33:** 591–601.

Zhou A, Ou AC, Cho A, Benz EJ, Jr., Huang SC. 2008. Novel splicing factor RBM25 modulates Bcl-x pre-mRNA 50 splice site selection. *Mol Cell Biol* **28:** 5924-5936.

CHAPTER 3: PQBP1, A FACTOR LINKED TO INTELLECTUAL DISABILITY, AFFECTS ALTERNATIVE SPLICING ASSOCIATED WITH NEURITE OUTGROWTH

ABSTRACT

Polyglutamine binding protein 1 (PQBP1) is a highly conserved protein associated with neurodegenerative disorders. Herein, we identify PQBP1 as an alternative mRNA splicing (AS) effector, capable of influencing splicing of multiple mRNA targets. PQBP1 is associated with many splicing factors including the key U2 snRNP component SF3B1. Loss of functional PQBP1 reduced SF3B1-substrate mRNA association and led to significant changes in AS patterns. Depletion of PQBP1 in primary mouse neurons reduced dendritic outgrowth and altered AS of mRNAs enriched for functions in neuron projection development. Disease-linked PQBP1 mutants were deficient in splicing factor associations and could not complement neurite outgrowth defects. Our results indicate that PQBP1 can affect AS of multiple mRNAs and indicate specific affected targets whose splice-site determination may contribute to the disease phenotype in PQBP1-linked neurological disorders.

INTRODUCTION

Alternative splicing (AS) is a major source for biological diversity and a crucial determinant of cell identity and fate. To date, more than 90% of human precursor messenger RNAs (pre-mRNAs) are reported to undergo AS, with immense variation across tissue types and developmental stages (Wang et al. 2008). The regulation of AS is carried out by the interaction of RNA-binding proteins (RBPs), notably hnRNPs and SR proteins, with cis-elements in the pre-mRNAs (Black 2003). The dynamics of RBP-cis-element recognition are further modulated by a wide spectrum of spliceosome-associated proteins and are responsive to signal transduction networks (Varani and Nagai 1998; Wahl et al. 2009). Tissue-specific AS regulation is especially prominent in the nervous system (Castle et al. 2008), where neurons must promptly respond to complex stimuli and perform downstream functions with flexibility and precision (Dredge et al. 2001). Accordingly, neuron-specific AS networks modulate such diverse functions including neurotransmission, receptor activity and ion channel function, and thus impact higher brain functions including cognition, coordination and learning (Grabowski and Black 2001).

Aberrant AS is a primary cause for many human diseases (Wang and Cooper 2007). Diverse neurological disorders are closely linked to aberrant splicing and can be classified as cis-acting and trans-acting splicing defects (Licatalosi and Darnell 2006). Cis-acting defects originate from mutations in the disease-associated gene that change sequences important for the correct splicing of the pre-mRNA. A representative example is frontotemporal dementia with Parkinsonism linked to chromosome 17, which arises from mutations causing mis-splicing of the mRNA for the microtubule-associated protein

tau (Licatalosi and Darnell 2006). Trans-acting defects arise from mutations or misregulation of splicing regulatory factors. One example is retinitis pigmentosa characterized by a progressive loss of rod photoreceptor cells, which results from mutations in HPRP3, PRP31, and PRPC8—factors crucial for the assembly and functioning of U snRNPs in pre-mRNA splicing (McKie et al. 2001; Vithana et al. 2001; Chakarova et al. 2002). Although increasing numbers of splicing-related disorders have been identified, the molecular link between AS regulation and the disease phenotype remains to be elucidated.

Polyglutamine binding protein 1 (PQBP1) is a highly conserved protein with mutations associated with about 10 familial X-linked mental retardation (XLMR) diseases (Kalscheuer et al. 2003; Stevenson et al. 2005). Patients with these diseases share similar symptoms such as intellectual deficiency, microcephaly, and short stature (Kalscheuer et al. 2003; Stevenson et al. 2005). Although these diseases were separately discovered and originally thought to be unrelated, they were recently grouped together as "Renpenning Syndrome" because of the common symptomatic manifestation and mutations in PQBP1 (Stevenson et al. 2005).

The molecular function of PQBP1 remains unclear. PQBP1 is linked to transcription regulation based on its interaction with RNA Polymerase II and other transcription factors (Waragai et al. 1999; Okazawa et al. 2002). Recent studies also reported that PQBP1 was dynamically associated with stress granules and interacted with dynactin components, indicating a possible role for PQBP1 in RNA transport or cytoplasmic RNA processing (Kunde et al. 2011). Also, PQBP1 has been implicated in pre-mRNA splicing: PQBP1 interacts with known splicing factors, such as WBP11 and

U5-15KD (Komuro et al. 1999; Waragai et al. 2000) and is part of the early spliceosome (Makarova et al. 2004). Consistent with this view, a disease-associated mutant of PQBP1 decreased splicing efficiency (Tapia et al. 2010).

We previously identified PQBP1 in a screen for factors that affect the AS of the mRNA for Bcl-X, a mammalian apoptosis regulator (Moore et al. 2010). Here, we show that PQBP1 affects AS of numerous other pre-mRNAs in human cells. In addition, we used whole-transcriptome sequencing to identify a global set of PQBP1 AS targets in primary mouse neurons, which revealed that PQBP1 coordinates important neuronal functions and maintains normal neurite development. Collectively, our findings indicate that aberrant AS contributes to the disease pathology of PQBP1-associated neurological disorders.

RESULTS

PQBP1 affects the AS of mRNAs encoding apoptotic factors

We previously identified PQBP1 as a potential effector for *Bcl-x* splicing in HeLa cells in a genome-wide siRNA screen using a fluorescence-based minigene assay (Moore et al. 2010; Figure 3-1A). To verify the effect of PQBP1 on the AS of *Bcl-x*, we examined the splicing of endogenous *Bcl-x* after depletion of PQBP1 with two different siRNAs (Figure 3-1B) using RT-PCR primers flanking the alternatively spliced region (Figure 3-1A). As expected, depletion of PQBP1 favored a shift from the anti-apoptotic isoform *Bcl-xL* toward the pro-apoptotic isoform *Bcl-xS*, via increased usage of an alternative 5' splice site in the second exon (Figure 3-1A,C; Supplemental Figure S3-1A,B). These results confirmed that PQBP1 affects the AS of *Bcl-x* in HeLa cells.

Figure 3-1: AS targets of PQBP1 in HeLa cells.

(A) Schematic of Bcl-x. Light rectangles are exons, with dark color marking the alternatively spliced region; black lines are introns. Arrows show translation start site; arrowheads mark qPCR primer sets.

(B) Western blot of HeLa cell lysates where PQBP1 is knocked down by siRNA. The control sample was treated with a siRNA targeting firefly luciferase mRNA. Antibodies are listed on the right. Knockdown with two different siRNAs is shown.

(C) Gel analysis of RT-PCR of Bcl-x splicing isoforms for control and PQBP1 KD samples is shown (left panel), with quantification (right panel). Mean of three independent measurements +/- SD are shown. Changes were tested by 1-way ANOVA. **: statistically significant with p-value<0.01. See also Supplemental Figure S3-1A,B.

(D) Quantification of real-time PCR data on 14 mRNAs that showed significant AS changes between control and PQBP1 KD samples, with Mcl1 (not a target of PQBP1) as the negative control. Y-axis depicts the ratio between amplicons for constitutive and alternative exonic regions; TPX-2 and CASP6 follow the y-axis on the right side while the rest follow the y-axis on the left side. Light columns show the ratio of two amplicons from the control sample and dark columns from the PQBP1 KD sample. Mean of three independent measurements +/- SD are shown. Difference of the ratio between control and PQBP1 KD samples was analyzed by 1-way ANOVA test. ∗: statistically significant with p-value<0.03; ∗∗: statistically significant with p-value<0.01; n.s.: not significant. See also Supplemental Figure S3-1C.

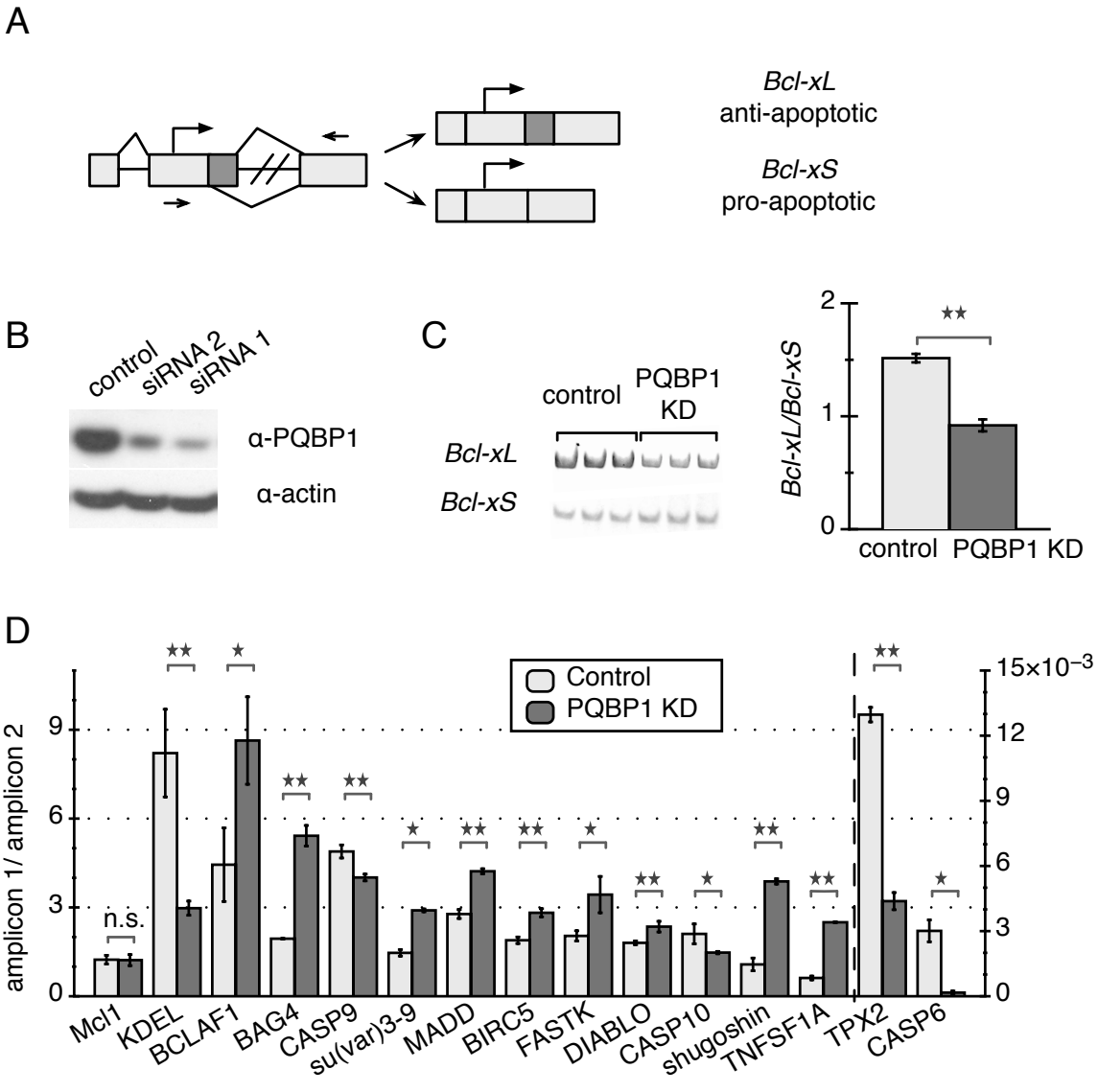

Figure 3-1 (Continued).

We now show that PQBP1 affects additional AS events. We analyzed 31 mRNAs encoding apoptosis-related factors with annotated alternatively spliced isoforms. We designed primers to amplify the alternatively spliced exonic region and constitutive exonic region of each mRNA and the ratio of the two amplicons was compared between the control and PQBP1 knockdown (KD) samples by RT-PCR. Approximately half of these mRNAs showed significant changes in AS when PQBP1 was knocked down (Figure 3-1D; Supplemental Figure S3-1C). These results provided further evidence that PQBP1 affects the AS of many mRNAs.

PQBP1 is associated with splicing factors

To further clarify PQBP1's role in AS regulation, we systematically determined the spectrum of proteins that are associated with PQBP1. In addition, to understand the aberrant function of disease-related PQBP1 mutants, we compared the associated protein profile of wildtype (WT) PQBP1 and two mutant variants identified in Renpenning Syndrome patients.

PQBP1 protein contains a WW domain, a polar amino-acid rich domain and a carboxyl-terminal domain with a nuclear localization signal (NLS) (Waragai et al. 1999; Figure 3-2A). The mutations in PQBP1 identified from Renpenning Syndrome patients can be classified into two categories: frame-shifting mutations in the AG dinucleotide hexamer in the polar amino-acid rich domain and missense mutations in the WW domain (Musante et al. 2009). Here, we examined two representative disease-linked variants—one with an AG di-nucleotide deletion (ΔAG) in the polar amino-acid rich domain resulting in a truncated protein and one with a point mutation (Y65C) in the WW domain

Figure 3-2: The spectrum of proteins associated with PQBP1 and disease-linked mutants.

(A) Schematics of the protein domain organization of WT PQBP1 and two disease-linked mutants. Boxes are protein domains with names shown in the center (WW: WW domain; NLS: nuclear localization signal; CTD: carboxyl-terminal domain). Red arrowheads show the mutation sites in the disease-linked PQBP1 variants.

(B) Immunofluorescence (IF) staining of HeLa cells stably expressing FLAG-HA tagged WT PQBP1, ΔAG or Y65C. Cells were fixed and stained with antibodies or dyes listed on the bottom of each panel column.

(C) Western blot of HeLa cell lines stably expressing empty FLAG-HA vector, FLAG-HA tagged WT PQBP1, FLAG-HA tagged ΔAG and FLAG-HA tagged Y65C. Antibodies used are indicated at the bottom of each panel.

(D) Silver staining of proteins co-purified with FLAG-HA tagged empty plasmid, FLAG-HA tagged WT PQBP1, FLAG-HA tagged ΔAG or FLAG-HA tagged Y65C.

(E) Gene Ontology (GO) term enrichments of proteins that were pulled down with WT PQBP1. Statistics were calculated with human genes as the background. X-axis is the corrected p-value (fdr) in negative log for enrichment.

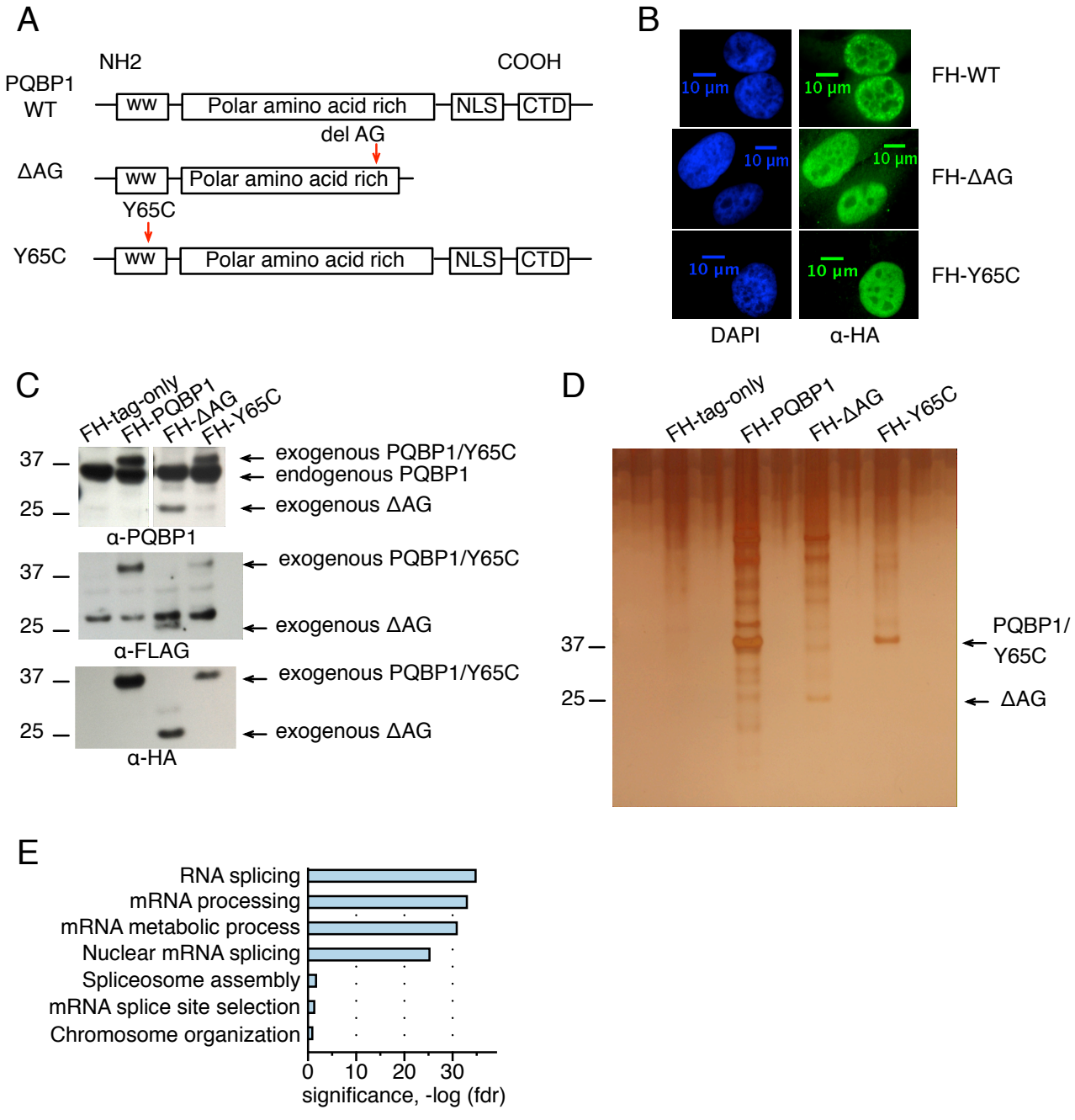

Figure 3-2 (Continued).

(Kalscheuer et al. 2003; Figure 3-2A). PQBP1 localizes predominantly in the nucleus (Waragai et al. 1999; Komuro et al. 1999; Kalscheuer et al. 2003; Figure 3-2B). Interestingly, both ΔAG and Y65C mutants localize to the nucleus (Figure 3-2B), although ΔAG loses a potential NLS (Figure 3-2A) indicating there must be an additional unidentified NLS or it enters together with another protein(s).

Proteins associated with WT PQBP1 were strongly enriched for known or potential splicing factors and spliceosome components (Figure 3-2E; Table 3-1). HeLa cells expressing stably integrated FLAG-HA tagged full-length PQBP1, ΔAG or Y65C were generated by stable integration of retroviral constructs, and shown by Western blotting to express exogenous proteins at a level comparable to endogenous PQBP1 (Figure 3-2C). Tandem affinity purification with FLAG and HA antibodies followed by mass spectroscopy was performed to identify PQBP1- or mutant-associated proteins (Figure 3-2D; Table 3-1). WBP11, a previously known PQBP1 interactor (Komuro et al. 1999) was detected in all three replicate purifications of WT PQBP1, confirming the validity of our approach (Table 3-1). U5-15KD, another proposed interactor of PQBP1 (Waragai et al. 2000), was not identified possibly because of cell specificity or the interaction is not strong enough to be detected in our tandem affinity purification. ΔAG showed significantly reduced splicing factor association compared to WT PQBP1, while Y65C lost associations with all splicing-related factors (Table 3-1). These results established an association, either direct or indirect, of PQBP1 to the splicing machinery and identified the WW domain as important for linking PQBP1 to splicing complexes. This finding is consistent with the idea that disease-linked mutations may disrupt splicing related functions of PQBP1.

Table 3-1. Comparison of associated protein spectrums for WT PQBP1, ΔAG and Y65C.

83 proteins identified to be associated with WT PQBP1, ΔAG or Y65C are listed (Y=associated with the corresponding PQBP1 or variants; N=not associated). Associated proteins for each PQBP1 protein were first pulled-down by tandem affinity IP and then identified from LC-MS/MS. The list was further filtered by deleting proteins that appeared also in the empty FLAG-HA vector control IP sample. The rest are listed, sorted by a general categorization based on their functions.

		WT PQBP1	ΔAG	Y65C
Protein Description	Entrez Gene Symbol			
WW domain-binding protein 11	WBP11	Y	Y	N
splicing factor 3B subunit 4	SF3B4	Y	N	N
splicing factor 3B subunit 3	SF3B3	Y	Y	N
splicing factor 3B subunit 2	SF3B2	Y	Y	N
splicing factor 3B subunit 1	SF3B1	Y	Y	N
splicing factor, suppressor of white-apricot homolog	SFSWAP	Y	Y	N
serine/arginine-rich splicing factor 7	SRSF7	Y	N	N
serine/arginine-rich splicing factor 6	SRSF6	Y	Y	N
serine/arginine-rich splicing factor 5	SRSF5	Y	N	N
serine/arginine-rich splicing factor 3	SRSF3	Y	Y	N
serine/arginine-rich splicing factor 2	SRSF2	Y	N	N
serine/arginine-rich splicing factor 1	SRSF1	Y	Y	N
serine/arginine repetitive matrix protein 2	SRRM2	Y	N	N
serine/arginine repetitive matrix protein 1	SRRM1	Y	Y	N
splicing factor 45	RBM17	Y	Y	N
RNA-binding protein 14	RBM14	Y	N	N
RNA-binding protein 10	RBM10	Y	Y	N
signal recognition particle 9 kDa protein	SRP9	Y	N	N
U2 snRNP-associated SURP motif-containing protein	SR140	Y	Y	N
U2 small nuclear ribonucleoprotein A'	SNRPA1	Y	N	N
survival of motor neuron-related-splicing factor 30	SMNDC1	Y	N	N
small nuclear ribonucleoprotein Sm D3	SNRPD3	Y	N	N
small nuclear ribonucleoprotein-associated proteins B and B'	SNRPB	Y	N	N
pre-mRNA-processing-splicing factor 8	PRPF8	Y	N	N

Table 3-1 (Continued).

pre-mRNA-processing factor 19	PRPF19	Y	N	N
heterogeneous nuclear ribonucleoprotein U-like protein 1	HNRNPUL1	Y	Y	N
heterogeneous nuclear ribonucleoprotein M	HNRNPM	Y	N	N
heterogeneous nuclear ribonucleoprotein F	HNRNPF	Y	N	N
heterogeneous nuclear ribonucleoprotein D	HNRNPD	Y	N	N
heterogeneous nuclear ribonucleoprotein A1	HNRNPA1	Y	N	Y
heterogeneous nuclear ribonucleoprotein A/B	HNRNPAB	Y	N	N
ATP-dependent RNA helicase A	DHX9	Y	Y	N
putative pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15	DHX15	Y	Y	N
probable ATP-dependent RNA helicase DDX5	DDX5	Y	Y	N
nucleolar RNA helicase 2	DDX21	Y	Y	N
probable ATP-dependent RNA helicase	DDX17	Y	Y	N
poly(U)-binding-splicing factor PUF60	PUF60	Y	N	N
polyadenylate-binding protein 4	PABPC4	Y	N	N
protein disulfide-isomerase precursor	P4HB	Y	N	N
putative RNA-binding protein Luc7-like 2	LUC7L2	Y	N	N
la-related protein 1	LARP1	Y	N	N
interleukin enhancer-binding factor 3	ILF3	Y	Y	N
interleukin enhancer-binding factor 2	ILF2	Y	N	N
insulin-like growth factor 2 mRNA-binding protein 3	IGF2BP3	Y	N	N
zinc finger CCCH-type antiviral protein 1	ZC3HAV1	Y	N	N
X-ray repair cross-complementing protein 6	XRCC6	Y	Y	N
X-ray repair cross-complementing protein 5	XRCC5	Y	Y	N
THO complex subunit 4	THOC4	Y	N	N
heterogeneous nuclear ribonucleoprotein Q	SYNCRIP	Y	N	N
activated RNA polymerase II transcriptional coactivator p15	SUB1	Y	Y	N
DNA-binding protein A	CSDA	Y	N	N
calcium homeostasis endoplasmic reticulum protein	CHERP	Y	Y	N
coiled-coil domain-containing protein 86	CCDC86	Y	N	N
hypothetical protein LOC84310	C7orf50	Y	Y	N
transcription factor IIB	C16orf80	Y	Y	N
high mobility group protein HMG-I/HMG-Y	HMGA1	Y	N	N
glutamate-rich WD repeat-containing protein 1	GRWD1	Y	N	N
guanine nucleotide-binding protein-like 3	GNL3	Y	N	N
ubiquitin-like protein fubi and ribosomal protein S30 precursor	FAU	Y	Y	N

Table 3-1 (Continued).

eukaryotic translation initiation factor 5B	EIF5B	Y	Y	N
DNA topoisomerase 2-alpha	TOP2A	Y	Y	N
DNA topoisomerase 2-beta	TOP2B	Y	N	N
DNA topoisomerase 1	TOP1	Y	Y	N
histone H2B type 1-C/E/F/G/I	HIST1H2BG	Y	Y	N
histone H2A type 1	HIST1H2AI	Y	N	N
histone H1.3	HIST1H1D	Y	Y	N
histone H1.5	HIST1H1B	Y	Y	N
histone H1x	H1FX	Y	N	N
replication factor C subunit 1	RFC1	Y	N	N
protein PRRC2A	PRRC2A	Y	N	N
serine/threonine-protein phosphatase PP1-beta catalytic subunit	PPP1CB	Y	N	N
prolyl 4-hydroxylase subunit alpha-1 isoform1 precursor	P4HA1	Y	Y	N
mRNA turnover protein 4 homolog	MRTO4	Y	N	N
MAP7 domain-containing protein 1	MAP7D1	Y	N	N
acyl-CoA dehydrogenase family member 11	ACAD11	Y	N	N
FACT complex subunit SSRP1	SSRP1	N	Y	N
SPATS2-like protein	SPATS2L	N	Y	N
polypyrimidine tract-binding protein 1	PTBP1	N	Y	N
histone H2A.Z	H2AFZ	N	Y	N
guanine nucleotide-binding protein G(t) subunit alpha-3	GNAT3	N	N	Y
acetyl-CoA acetyltransferase, mitochondrial precursor	ACAT1	N	N	Y
GTP-binding nuclear protein Ran	RAN	N	N	Y
FERM domain-containing protein 5	FRMD5	N	N	Y

PQBP1 influences SF3B1's recognition of splicing sites

To explore the mechanism of PQBP1 in AS regulation, we first examined whether PQBP1 directly binds pre-mRNAs whose splicing it affects. Although previous studies showed PQBP1 bound to polyrG resin *in vitro* (Komuro et al. 1999), PQBP1 protein itself has no known RNA-binding domains (Figure 3-2A; Lunde et al. 2007). We used the CLIP (cross-linking and immunoprecipitation) method to determine if PQBP1 directly binds its targets *in vivo* (Ule et al. 2003). Live HeLa cells were exposed to UV light to crosslink RNA-protein complexes, and then PQBP1 was immunopurified from cell lysates (Figure 3-3A). RT-PCR quantification of *Bcl-x* and 9 other target mRNAs (Figure 3-1D) showed no enrichment in PQBP1 IPs compared to an irrelevant IgG control. These results suggest that PQBP1 does not bind AS target mRNAs directly with the caveat that it may fail to be effectively cross-linked under the conditions where other known splicing factors are observed to bind.

PQBP1, on the other hand, affects the association of the spliceosome factor SF3B1 with a target pre-mRNA. Among the proteins that are associated with PQBP1 are the splicing factor 3B (SF3B) protein complex members (Table 3-1), which form part of the U2 small nuclear ribonucleoprotein (U2 snRNP). The function of the SF3B complex is critical for the assembly of the spliceosome complex A as well as the successful execution of the splicing reaction (Wahl et al. 2009). Furthermore, subunit 1 of the SF3B complex—SF3B1—was reported to regulate the AS of *Bcl-x* through direct interaction with a cis-element in the *Bcl-x* mRNA (Massiello et al. 2006). And we observed that CLIP with SF3B1 showed a 40 fold enrichment of *Bcl-x* binding over an IgG control (Figure 3-3D, blue column over *Bcl-x*).

Figure 3-3: SF3B1 function is affected by PQBP1.

(A) Western blot of PQBP1 CLIP together with an irrelevant IgG control, probed with PQBP1 antibody. PQBP1 was not detected in depleted supernatants or the IgG pulldown, indicating efficient and specific antibody binding of PQBP1.

(B) Western blot of SF3B1 CLIP from PQBP1 KD samples and samples treated with siRNA targeting firefly luciferase mRNA as control. Shown are total cell lysates of control (lane 1) and PQBP1 KD (lane 2) HeLa cells probed with antibodies indicated on the left side.

(C) Western blot of SF3B1 CLIP from PQBP1 KD samples and samples treated with siRNA targeting firefly luciferase mRNA as control. Lane 1 and 2 are the IgG controls; Lane 3 and 4 are SF3B1 CLIP stained by SF3B1 antibody for control and PQBP1 KD samples.

(D) Quantification of *Bcl-x* and *Mcl1* mRNA enrichment in SF3B1-RNA complex as from SF3B1 CLIP in control and PQBP1 KD samples respectively by RT-PCR. Data values normalized to the IgG control and mean of three independent measurements +/- SD are shown.

(E) Proposed model mechanism of PQBP1 AS regulation. A schematic of part of a hypothetical mRNA is shown. Orange rectangles are exons and the black line is the intron. “GU” marks the 5’ splice site, “A” the branch site recognized by the U2 snRNP with help from components like SF3B1 and “AG” the 3’ splice site. Dashed double green lines indicate the association between PQBP1 and SF3B1. Through an association with SF3B1 and other splicing factors, PQBP1 influences the recruitment of U2 snRNP to specific target sites, thus affecting splicing decisions on a subset of mRNAs. Loss or mutations of PQBP1 interrupt the splicing factor association and the recognition of specific splice sites. See Discussion.

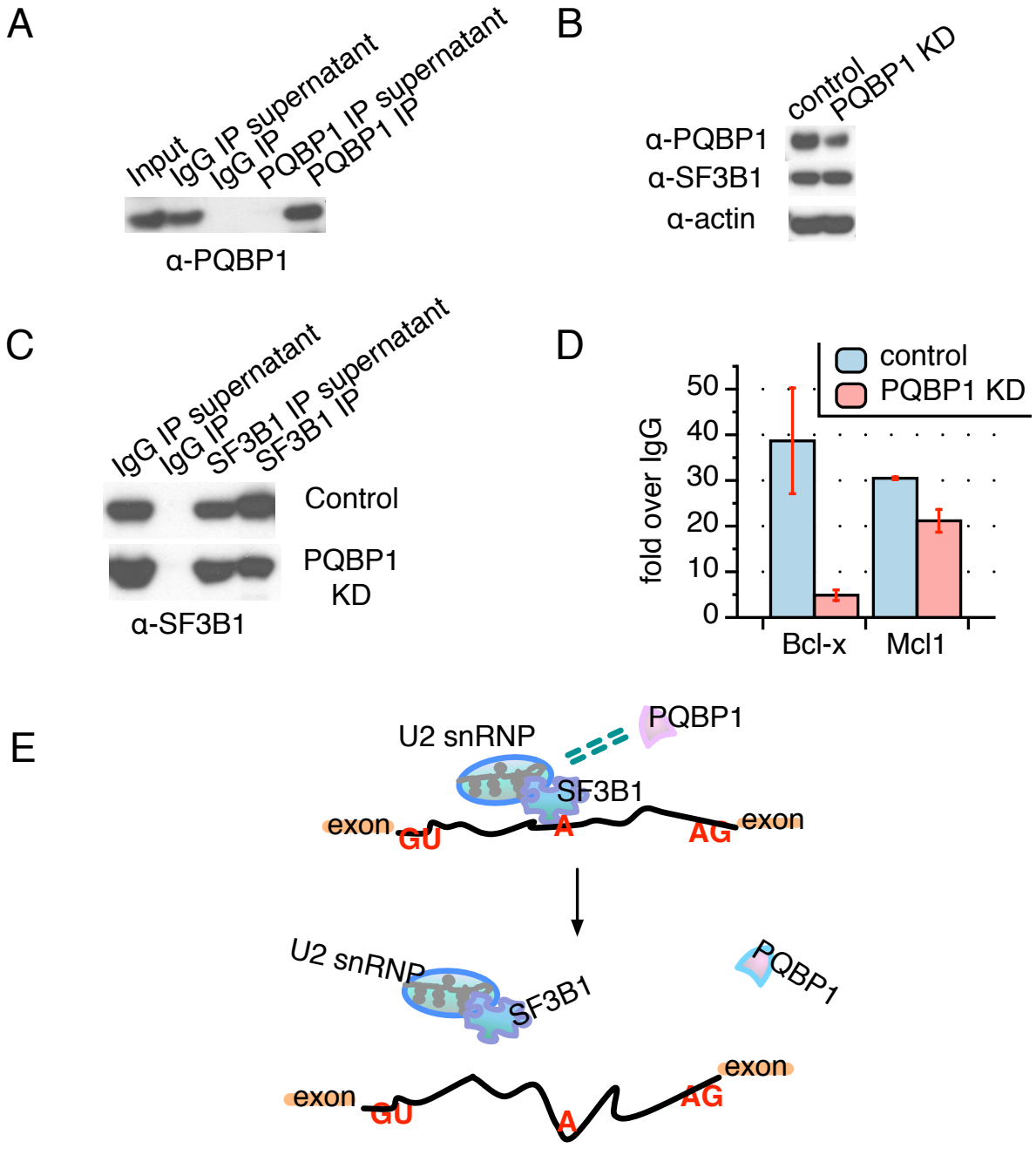

Figure 3-3 (Continued).

To test if the recognition of *Bcl-x* by SF3B1 is influenced by PQBP1, we knocked down PQBP1 in HeLa cells and re-analyzed the enrichment of *Bcl-x* in the SF3B1-RNA complex (Figure 3-3B,C). PQBP1 depletion led to an 8-fold decrease of *Bcl-x* mRNA levels in SF3B1 CLIP compared to a control knockdown (Figure 3-3D), although the level of SF3B1 protein remained unchanged (Figure 3-3B) and the steady state level of *Bcl-x* mRNA did not change upon PQBP1 KD (Supplemental Figure S3-2).

To further test if PQBP1's impact on SF3B1-RNA recognition is specific to the targets of PQBP1, we analyzed the mRNA of *Mcl1* in the SF3B1-RNA complex. Like Bcl-X, Mcl1 is a mammalian apoptosis regulator and has two splicing isoforms, anti-apoptotic *Mcl1-L* and pro-apoptotic *Mcl1-S* (Akgul et al. 2004). However, while knockdown of SF3B1 significantly shifted the splicing of *Mcl1* towards *Mcl1-S* (Moore et al. 2010), PQBP1 knockdown had little affect (Figure 3-1D; Moore et al. 2010). We therefore hypothesized that the knockdown of PQBP1 should have relatively minor influence on *Mcl1* mRNA enrichment in the SF3B1-RNA complex. Indeed, CLIP data showed only a 1.4 fold decrease of *Mcl1* mRNA enrichment in the SF3B1-RNA complex in PQBP1 KD sample compared to control (Figure 3-3D). Additionally, the decrease was mostly likely due to *Mcl1* mRNA steady state level change of 1.5-fold (Supplemental Figure S3-2). Therefore, PQBP1 influenced the recognition of SF3B1 to one mRNA whose AS is affected by PQBP1, but not for an mRNA where AS was regulated by SF3B1 but not PQBP1. These results do not rule out possibilities that PQBP1 may affect mRNA AS through other interactions, or that other splicing factors may assist in the process.

Loss of PQBP1 causes defects in neurite outgrowth

Previous work found that PQBP1 is most highly expressed in the central nervous system of embryonic or newborn rodents, with the peak around birth (Qi et al. 2005). Considering the tissue-specific expression pattern of PQBP1 as well as its disease association, we examined the role of PQBP1 in AS modulation in embryonic rodent neurons.

PQBP1 is highly expressed in embryonic mouse cortical and hippocampal neurons as shown by Western blots (Figure 3-4A). In addition, PQBP1 localizes predominantly in the nucleus of neurons (Figure 3-4B, upper panels), as shown previously in other cell types (Waragai et al. 1999; Komuro et al. 1999; Kalscheuer et al. 2003). A closer examination demonstrated that PQBP1 co-localizes with SC35—a splicing factor in nuclear speckles, a nuclear domain enriched for splicing regulatory factors (Figure 3-4B, lower panels; Supplemental Figure S3-3A; Spector and Lamond, 2011).

To understand the importance of PQBP1 in neuron development, we knocked down PQBP1 in mouse embryonic primary cortical neurons with shRNAs delivered by lentivirus (Figure 3-4C). Neurons were first examined by immunostaining of the microtubule-associated protein 2 (MAP2), a marker for dendrites. Interestingly, upon PQBP1 KD the number of projections and length of neurites were decreased and the connection between neurons degenerated considerably (Figure 3-4D). A second shRNA against PQBP1 led to a similar defect in neurite morphology (Supplemental Figure S3-3B,C). To determine if this defect is simply due to cell apoptosis, we compared the number of DAPI-stained intact nuclei after fixation as well as the percentage of

Figure 3-4. PQBP1 affects dendritic outgrowth and branching in mouse cortical neurons.

(A) Western blot of PQBP1 expression in mouse embryonic cortical and hippocampal neurons, compared to endogenous β-actin expression.

(B) IF staining of mouse embryonic primary cortical neurons (the four upper panels) and rat embryonic primary cortical neurons (the three lower panels). Neurons were fixed and stained with antibodies or dyes listed on top of each panel. See also Figure S3-3A.

(C) Western blot of mouse embryonic cortical neurons infected with virus from a non-targeting control shRNA or a PQBP1-targeting shRNA. See also Figure S3-3B.

(D) IF staining of mouse primary embryonic cortical neurons 5 days after infection with virus from a non-targeting control shRNA or a PQBP1-targeting shRNA. Green: MAP2 staining; blue: DAPI staining. See also Figure S3-3C.

(E) Comparison of the number of intact nuclei between neuron samples 5 days after infection with virus from a non-targeting control shRNA or a PQBP1-targeting shRNA. Neurons were fixed and stained with DAPI. Healthy and intact nuclei were counted for each sample. Mean of eight independent measurements +/- SD are shown. Difference in the number of intact nuclei between control and PQBP1 KD samples was analyzed by 1-way ANOVA test. n.s.: not significant

(F) Comparison of the percentage of Annexin-V–stained neurons between samples infected with virus from a non-targeting control shRNA or a PQBP1-targeting shRNA. The number of neurons with Annexin V staining and the percentage of Annexin-V–positive neurons were calculated and normalized to neuron samples that were not treated by virus. Y-axis depicts the fold of the percentage of Annexin-V–positive neurons in control or PQBP1 KD samples to samples that were untreated. Neurons on day 3 and day 6 after infection/plating were studied. Mean of six independent measurements +/- SD are shown. Difference in the percentage of Annexin-V–positive neurons between control and PQBP1 KD samples was analyzed by 1-way ANOVA test. n.s.: not significant

(G) IF of individual neurons that were transiently transfected with either a non-targeting control shRNA or a PQBP1-targeting shRNA together with a GFP-vector. Neurons were grown on a layer of feeder glial cells and were examined 5 days after transfection.

(H) Sholl analysis of dendritic branching for neurons transiently transfected with a non-targeting control shRNA or PQBP1-targeting shRNA together with a GFP-vector. Concentric circles were drawn at intervals of 5 μm. The number of dendritic branch intersections with each concentric

circle was counted (mean ± S.E., n=16-17). The x-axis depicts the distance of the concentric circle from the neuron soma center and the y-axis depicts the number of intersections the corresponding circle encounters with dendrites.

Figure 3-4 (Continued).

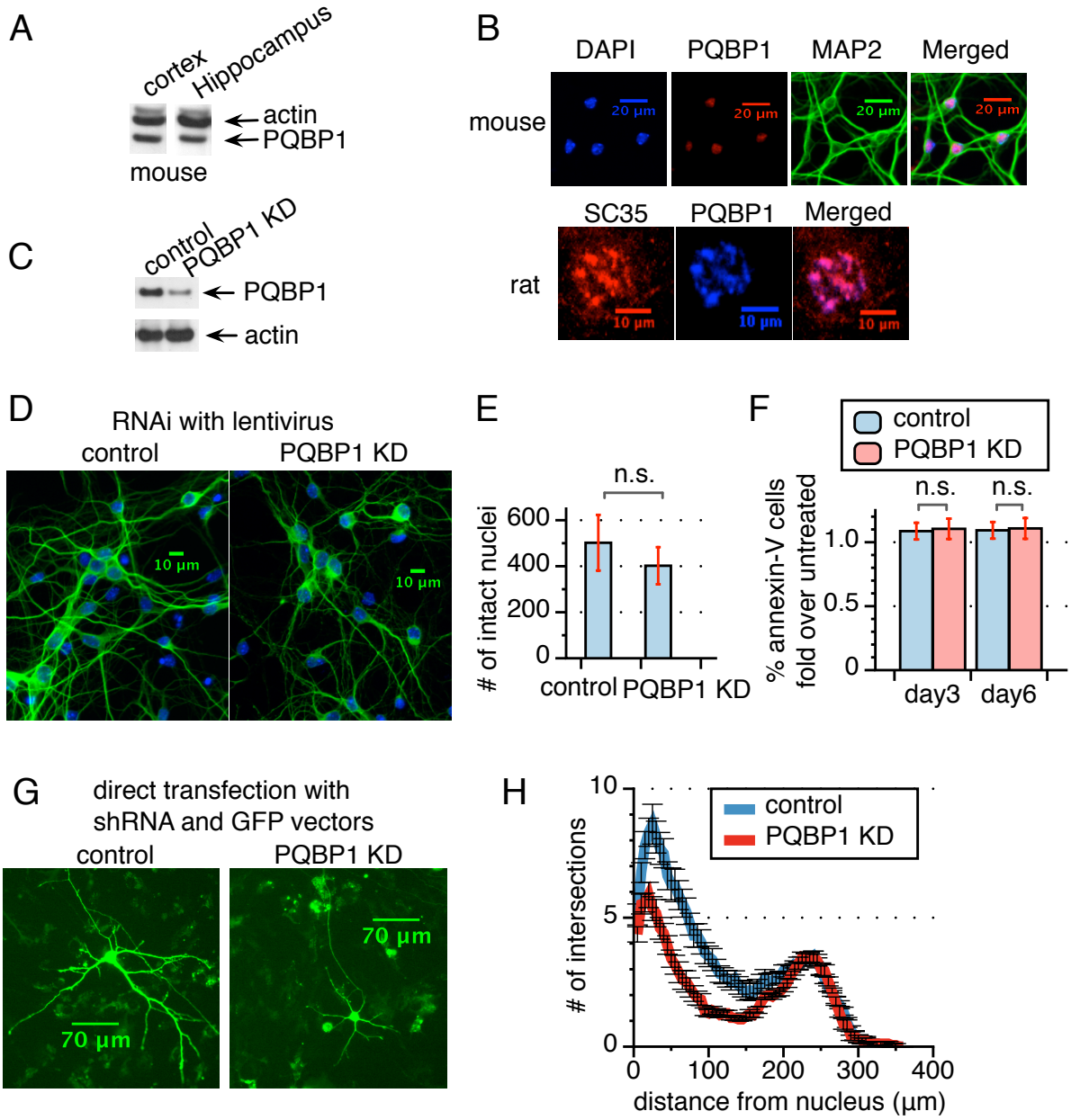

Figure 3-4 (Continued).

Annexin-V–stained cells between control and PQBP1 KD neurons (Figure 3-4E,F). No significant differences were observed, indicating that the defects observed from PQBP1 KD are not from neuron apoptosis.

PQBP1 KD neurons showed substantially diminished dendrite branching and dendrite length (Figure 3-4G,H). To quantify the changes in neurite outgrowth, we transiently transfected neurons with a non-targeting control shRNA or a PQBP1-targeting shRNA together with a GFP vector (Figure 3-4G). Sholl analysis was then performed on neurons that were GFP-highlighted for control and PQBP1 KD (Figure 3-4H; Sholl, 1953).

To confirm that the dendritic outgrowth defect observed was a direct effect of PQBP1 KD, we examined whether the defect could be rescued by re-expressing WT PQBP1 in comparison to the disease-associated PQBP1 mutants. shRNA-resistant, FLAG-HA-tagged cDNAs encoding human PQBP1, ΔAG and Y65C were co-expressed in mouse cortical neurons with shRNAs targeting endogenous mouse PQBP1 (Figure 3-5A; Supplemental Figure S3-3D). Western blots confirmed the knockdown of endogenous mouse PQBP1 and the concomitant expression of exogenous dual-tagged human PQBP1 and variants (Figure 3-5B). Total PQBP1 expression in all three rescue samples was compatible with the level of PQBP1 in the control sample, with the exception of ΔAG, which reached only 68% of the control (Figure 3-5B). To quantitatively compare neurite outgrowth between mutant PQBP1 and WT rescue, we transfected mouse cortical neurons with the rescue constructs (Figure 3-5C) and performed Sholl analysis after five days. A significant restoration of dendrite density and length was observed in WT PQBP1 rescued neurons, while ΔAG and Y65C showed less

improvement (Figure 3-5D,E,F). In the case of ΔAG, it is possible that insufficient expression of ΔAG may contribute to the deficiency of rescue.

Combining the results above, we conclude that PQBP1 is important in maintaining neurite projection and outgrowth in neuron development. Moreover, deviation in neurite structure may be part of the defects brought about by the disease-associated mutants of PQBP1.

Coupling of neuron projection and PQBP1 function

Dendrite structure is the crucial basis for neuron network formation and information processing. To clarify the molecular function of PQBP1 in maintaining neurite outgrowth, we systematically profiled the AS targets of PQBP1 in mouse embryonic cortical neurons by RNA-seq.

We considered RNA-seq reads that were directly mapped over splice junctions as the most straightforward evidence to quantify the splicing changes. For each splice junction, a 5' initiation site (5' IS) is defined as the final nucleotide position of the upstream exon, and the 3' end site (3' ES) is defined as the first nucleotide position of the downstream exon (Figure 3-6A). An AS event is then defined as a set of junctions that share the same 5' IS or the same 3' ES, with each junction in the set referred to as a sub-AS-junction (Figure 3-6A). This definition covers all 6 primary AS patterns except for intron retention, which was not considered in our analysis (Figure 4-2; Black 2003). An AS event under an experimental condition is represented by the distribution of read counts mapped to each sub-AS-junction within the AS event—i.e., the 'usage' of each sub-AS-junction under that condition. We developed methods to identify AS events that

Figure 3-5: WT PQBP1 but not the mutants rescues the neuron morphological defects

(A) Schematic of the pLKO.1-derived lentiviral constructs used to generate viruses for the rescue experiments. See also Figure S3-3D.

(B) Western blot of neuron samples that were infected with virus from an unmodified PQBP1-targeting shRNA, a non-targeting control shRNA, or PQBP1-targeting shRNA constructs co-expressing exogenous WT PQBP1, ΔAG or Y65C. Antibodies used are listed on the right. Quantification of total PQBP1 level for each sample compared to control is shown below each lane.

(C) IF of individual neurons that were transiently transfected with PQBP1-targeting shRNA constructs co-expressing exogenous WT PQBP1, ΔAG or Y65C together with a GFP-vector. Neurons were grown on a layer of feeder glial cells.

(D-F) Sholl analysis of dendritic branching for neurons that were transiently transfected with a non-targeting control shRNA, an unmodified PQBP1-targeting shRNA, or PQBP1-targeting shRNA constructs co-expressing WT PQBP1 (D), ΔAG (E) or Y65C (F) (mean ± S.E., n=10-17).

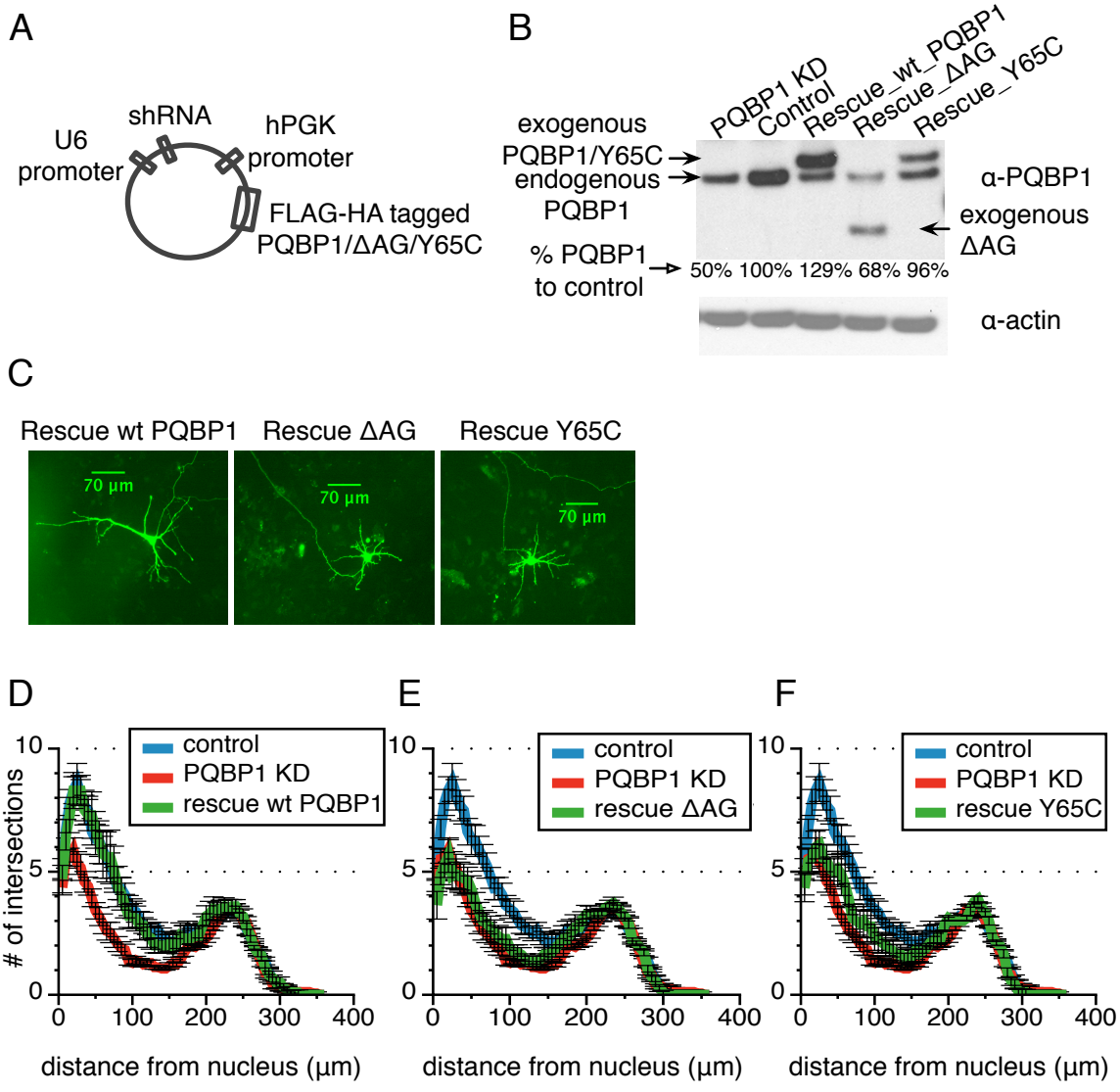

Figure 3-5 (Continued).

experienced significant differential 'usage' of any sub-AS-junction between control and PQBP1 KD conditions (for details, see Chapter 4), and defined these AS events as AS targets of PQBP1. In total, 457 targets were identified in the process (Figure 4-3; Supplemental Table S4-1).

To validate PQBP1 targets identified by RNA-seq, we picked 10 candidate targets together with two candidate non-targets from the analysis and designed primers to amplify each splicing isoform of each gene. We then compared the ratio of the isoform expression levels by RT-PCR for control and PQBP1 KD samples. All 10 identified targets exhibited significant AS shift upon PQBP1 KD while the two non-targets showed no discernible changes (Figure 3-6B). This result demonstrated the sensitivity and specificity of our computational method to identify AS targets of PQBP1.

The PQBP1 AS targets showed tissue-specific functional enrichment, with strong Gene Ontology (GO) enrichments for neuron projection development/morphogenesis, dendrite development and axonogenesis (Figure 3-6C; Supplemental Table S3-1), versus genes that were expressed in mouse embryonic cortical neurons profiled from the RNA-seq data. This observation was in accordance with the neurite outgrowth defect observed when PQBP1 was knocked down (Figure 3-4D,G,H). Targets of PQBP1 also showed strong enrichment in RNA splicing, synaptic transmission, chromatin modification, cell-cell signaling, phosphate metabolic process, neurotransmitter transport and ADP-ribosylation factor (ARF) protein signal transduction (Figure 3-6C; Supplemental Table S3-1). The GO term enrichment profile for PQBP1's AS targets is specific compared to high-throughput studies on other proteins in neurons, especially RNA splicing, chromatin modification, dendrite development related functions, and ARF signal transduction (Ule

Figure 3-6: PQBP1's AS targets in mouse embryonic cortical neurons.

(A) A schematic of a defined AS event and quantification of splicing changes. A hypothetical alternatively spliced mRNA is shown, with light rectangles representing exons and black lines between exons representing introns. The middle exon is the alternatively spliced exon. 5' IS is the start of a splicing junction and 3' ES the end of the junction. The AS event shown here includes two sub-AS-junctions that share the same 5' IS. Light dashed short lines mark RNA-seq reads that are mapped to sub-AS-junction-1 and dark solid short lines mark reads mapped to sub-AS-junction-2. The distribution of reads mapped to the two sub-AS-junctions was used for AS quantification in our analysis. See also Chapter 4.

(B) RT-PCR validation of 10 randomly picked AS events identified as PQBP1 targets and 2 randomly picked AS events identified as non-PQBP1-targets in the computational analysis. *Abat1* and *Icmt* were predicted non–PQBP1-targets. The y-axis depicts the ratio between isoform 1 and isoform 2. *Dlgap4*, *Tmod3, Adam15* and *Dclk1* follow the y-axis on the right side while the rest follow the y-axis on the left side. Mean of three independent measurements +/- SD are shown. Differences between control and PQBP1 KD samples were analyzed by 1-way ANOVA test. ∗: statistically significant with p-value<0.05; ∗∗: statistically significant with p-value<0.01; n.s.: not significant.

(C) GO term functional enrichment of AS targets of PQBP1. Statistics were calculated with all genes that are expressed in mouse embryonic cortical neurons (profiled from RNA-seq data) as the background. The GO term enrichment profile here is specific compared to RNA-seq studies for other proteins in neurons (see text). X-axis is the corrected p-value (fdr) in negative log for enrichment. See also Supplemental Table S3-1.

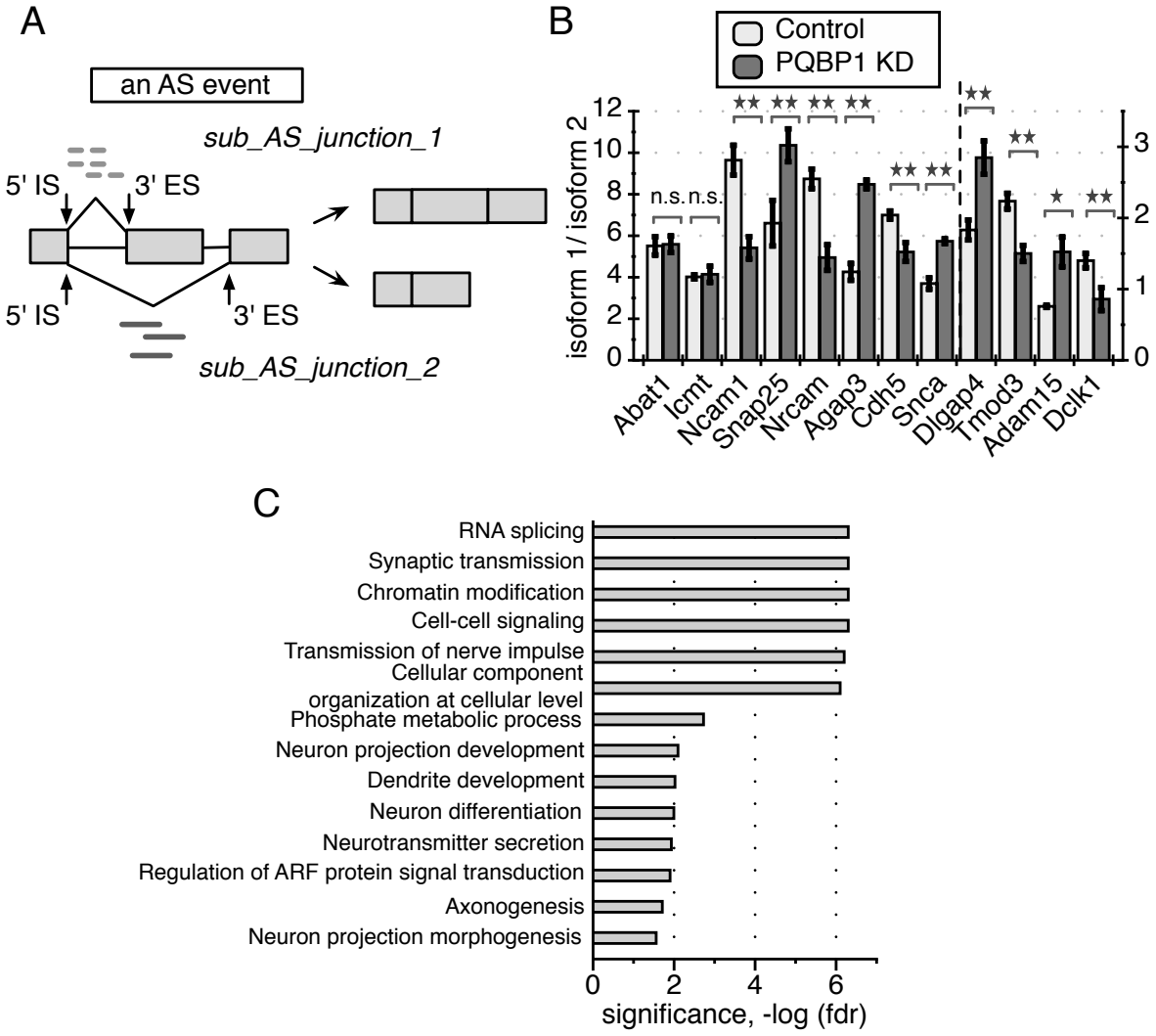

Figure 3-6 (Continued).

et al. 2005; Gehman et al. 2011; Charizanis et al. 2012). *Bcl-x* AS is not conserved from human to mouse, and was not identified as a target in this analysis. Among the 14 AS targets identified in HeLa cells that were related to apoptosis control (Figure 3-1D), only the mRNA of *BCLAF1* was recognized as an AS target of PQBP1 in mouse neurons, indicating the tissue specificity of PQBP1's regulation in AS.

Considering the physical and potential functional association of PQBP1 and SF3B1, we also knocked down SF3B1 in mouse cortical neurons and examined the AS of the 10 verified targets of PQBP1 (Supplemental Figure S3-4A). Interestingly, all 10 targets also showed a significant AS shift upon SF3B1 KD (Supplemental Figure S3-4).

Among the top hits of PQBP1 AS targets was NCAM-140, a major isoform of Neural Cell Adhesion Molecules 1 (Ncam1) (Figure 3-6B). Specifically, a 30nt-exon between exon 7 and 8 is alternatively spliced in NCAM-140 and is designated VASE (Variable Alternatively Spliced Exon) (Figure 3-7A). The VASE-excluding isoform of NCAM-140 (VASE –) promotes neurite outgrowth while the VASE-including isoform (VASE +) inhibits neurite outgrowth (Doherty et al. 1992). The embryonic brain mainly expresses VASE – with gradual ascendancy until final dominance of VASE + expression toward adulthood, except in regions such as the olfactory bulb where reformation of synapses and turnover of neurons continue (Small and Akeson 1990). Thus, the AS of VASE in NCAM-140 is a critical switch for neuronal plasticity. To confirm PQBP1's affect on VASE, we designed primers flanking the VASE exon and examined VASE + and VASE – isoforms by RT-PCR and electrophoresis (Figure 3-7A). VASE + was increased upon PQBP1 KD, while VASE – decreased (Figure 3-7B), resulting in a significant decrease in the ratio of VASE –/VASE + expression in response to PQBP1

shRNA versus a control hairpin (Figure 3-6B, *Ncam1*). These results were confirmed with a second shRNA targeting PQBP1 (Supplemental Figure S3-5).

As a further confirmation that PQBP1 impacts the AS of NCAM-140, we examined AS of NCAM-140 after PQBP1 KD in neurons co-expressing the WT PQBP1, ΔAG and Y65C constructs from Fig. 5A. Interestingly, neurons rescued with WT PQBP1 restored the ratio of VASE –/VASE + to levels observed in control samples (Figure 3-7C). In contrast, ΔAG and Y65C were less effective in correcting the VASE –/VASE + ratio (Figure 3-7C).

To test whether NCAM-140 is one of the PQBP1 targets responsible for the defects observed in cells lacking PQBP1, we examined the effect of restoring the ratio of VASE – and VASE +. We co-expressed VASE – under the hPGK promoter in the lentiviral shRNA vector as described above (Figure 3-5A). Knockdown of PQBP1 was verified by western blot (Figure 3-7D) and the change in the ratio of VASE –/VASE + by RT-PCR (Figure 3-7E). Neurons were transfected with the rescue construct (Figure 3-7F) and Sholl analysis on transfected neurons displayed a significant recovery from the neurite outgrowth defect caused by PQBP1 KD (Figure 3-7G). These results confirm the role of PQBP1 in the AS of NCAM-140 and modulating neurite outgrowth.

DISCUSSION

In this work, we have shown that the disease-associated factor PQBP1 plays a role in the AS of mRNAs. We demonstrated that PQBP1 associates with core spliceosome components and has a protein association network enriched for other splicing factors. Following downregulation of PQBP1, we observed aberrant splicing

Figure 3-7: PQBP1-modulated AS of NCAM-140 influences neurite outgrowth.

(A) A schematic of alternatively spliced region of NCAM-140. Green rectangles are exons and black lines are introns. The gray rectangle marks the Variable Alternatively Spliced Exon (VASE). Arrowheads are primers designed for RT-PCR and electrophoresis.

(B) NCAM-140 VASE splicing isoforms for neuron samples infected with virus from a non-targeting control shRNA or a PQBP1-targeting shRNA were analyzed by RT-PCR and gel electrophoresis. See also Supplemental Figure S3-5.

(C) RT-PCR quantification for VASE – and VASE +. The ratio of the two isoform levels are compared among neurons infected with virus from a non-targeting control shRNA, an unmodified PQBP1-targeting shRNA, or PQBP1-targeting shRNA constructs co-expressing exogenous WT PQBP1, ΔAG or Y65C. Mean of three independent measurements +/- SD are shown. Differences between samples were analyzed by 1-way ANOVA test. **: statistically significant with p-value<0.01

(D) Western blot of neurons infected with virus from a non-targeting shRNA, an unmodified PQBP1-targeting shRNA, a PQBP1-targeting shRNA construct co-expressing exogenous WT PQBP1, or a PQBP1-targeting shRNA construct co-expressing exogenous VASE – to restore the ratio of VASE –/VASE +.

(E) RT-PCR quantification for VASE – and VASE + expression levels of neurons infected with virus from a non-targeting shRNA, an unmodified PQBP1-targeting shRNA, or a PQBP1-targeting shRNA construct co-expressing exogenous VASE – to restore the ratio of VASE –/VASE +. Mean of three independent measurements +/- SD are shown. Differences between samples were analyzed by 1-way ANOVA test. **: statistically significant with p-value<0.01.

(F) IF of individual neurons that were transiently transfected with a PQBP1-targeting shRNA construct co-expressing exogenous VASE – to restore the ratio of VASE –/VASE + together with a GFP vector. Neurons were grown on a layer of feeder glial cells.

(G) Sholl analysis of dendritic branching for individual neurons that were transiently transfected with a non-targeting control shRNA, an unmodified PQBP1-targeting shRNA, or a PQBP1-targeting shRNA construct co-expressing exogenous VASE – to restore the ratio of VASE –/VASE + (mean ± S.E., n=15-17).

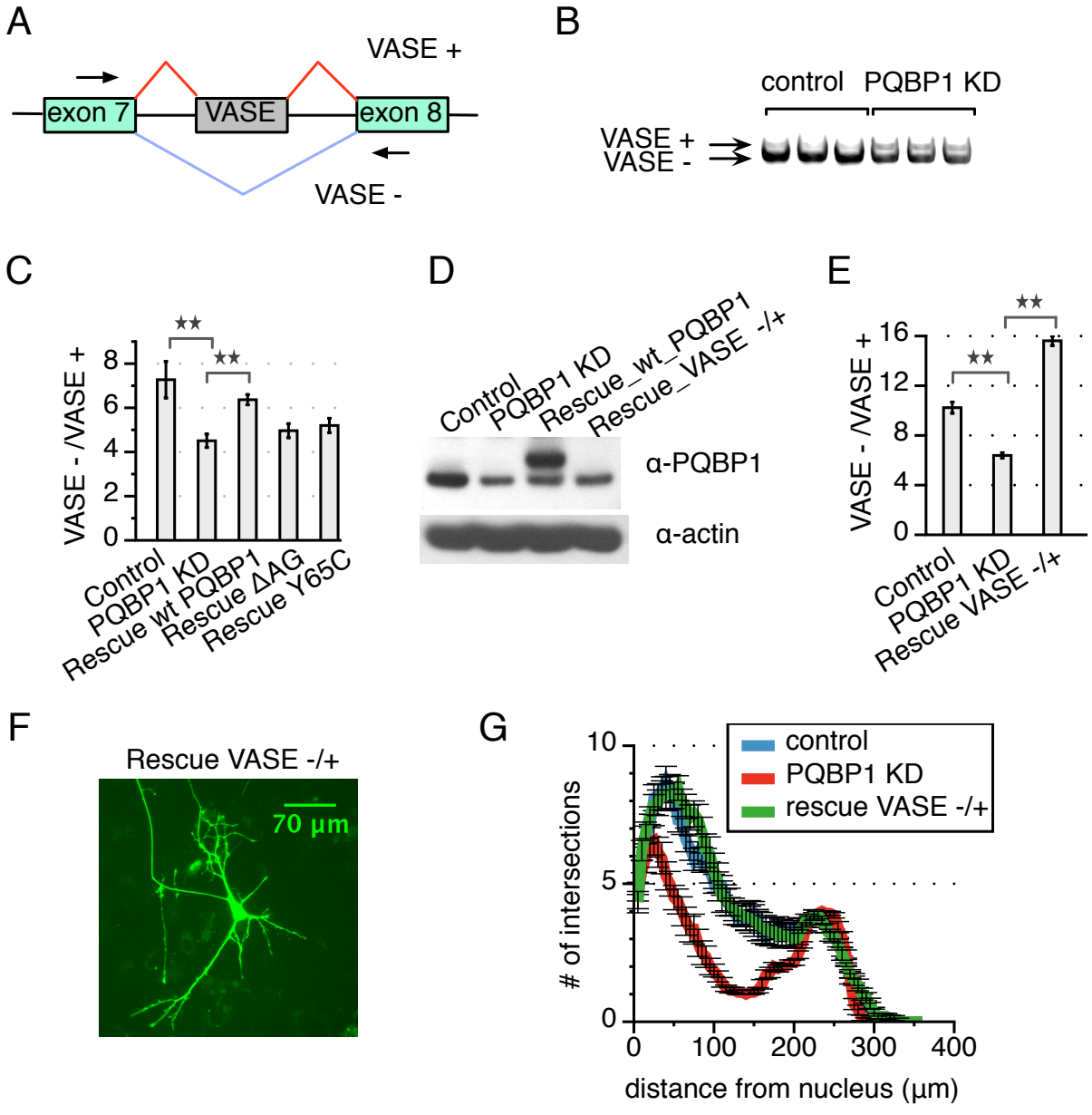

Figure 3-7 (Continued).

profiles for many mRNAs in both human cells and primary mouse cortical neurons. Specifically in neurons, mRNA profiling showed preferential misregulation of AS in factors that are important for neurite outgrowth, which is defective upon PQBP1 depletion. Collectively, we show that PQBP1 has a role in AS, and indicates aberrant AS as a possible mechanism underlying the symptoms of neurological disorders associated with PQBP1.

PQBP1 affects alternative splicing

Our work presents several lines of evidence that PQBP1 affects AS of many pre-mRNAs. Previous work has indicated an association of PQBP1 with mRNA processing factors (Waragai et al. 1999; Okazawa et al. 2002; Komuro et al. 1999; Waragai et al. 2000). We discovered that PQBP1 co-localizes with SC35 in nuclear speckles (Figure 3-4B; Supplemental Figure S3-3A), which are nuclear subdomains enriched for many splicing regulatory factors (Spector and Lamond 2011). Moreover, profiling of PQBP1-associated proteins revealed a set of factors highly enriched in splicing regulators and spliceosome components (Figure 3-2E; Table 3-1). Among these binding partners, we also identified WBP11, a previously reported interactor of PQBP1 (Table 3-1; Komuro et al. 1999). These results support PQBP1's connection with the splicing machinery.

PQBP1 affects U2 snRNP component SF3B1 function

We observed that PQBP1 did not associate with functional AS target mRNAs *in vivo* in CLIP experiments suggesting it does not bind directly to the AS targets identified in HeLa cells (Figure 3-3A) consistent with its apparent lack of known RNA-binding

domains (Figure 3-2A). SF3B1, which was previously shown to directly bind substrate mRNAs and affect *Bcl-x* AS (Gozani et al. 1998; Massiello et al. 2006) and was herein identified in PQBP1 affinity purifications (Table 3-1), constituted a prime candidate to work together with PQBP1. Consistent with this hypothesis, depletion of PQBP1 decreased SF3B1's association with *Bcl-x* mRNA (Figure 3-3D). In contrast, the binding of SF3B1 to *Mcl1*, a target of SF3B1 (but not of PQBP1) was not affected by the knockdown of PQBP1 (Figure 3-3D). These findings suggest that PQBP1 affects the association of SF3B1 with a distinct subset of substrate mRNAs. Further, we found that a subset of PQBP1 targets affected in their splicing was similarly influenced by the knockdown of SF3B1 (Supplemental Figure S3-4). SF3B1 is a central component of the U2 snRNP. Recruitment of U2 snRNP to the intronic branchpoint sequence (BPS) is a key step in the assembly of the spliceosomal A complex, and thus the selection of splice sites (Wahl et al. 2009). We propose a model whereby through an association with SF3B1, PQBP1 influences the recruitment of U2 snRNP to specific target sites, thus affecting splicing decisions on a subset of mRNAs (Figure 3-3E). The precise mechanism by which this occurs remains to be explored.

PQBP1 affects AS of mRNAs in neurons

Depletion of PQBP1 causes a decrease in outgrowth and branching of dendrites in mouse primary cortical neurons (Figure 3-4G,H; Supplemental Figure S3-3B,C). This morphological defect was not apoptosis-related (Figure 3-4E,F). Qualitatively, dendrites close to the nucleus appeared more affected than distal dendrites by knockdown of PQBP1 (Figure 3-4H; 3-5D). The defect in neurite morphology caused by loss of PQBP1

could be rescued by re-expression of WT PQBP1 (Figure 3-5C,D), confirming the correlation between neurite deformity and the loss of PQBP1 expression. Transcriptome-wide profiling revealed that PQBP1's AS targets were enriched for factors regulating neuron projection development (Figure 3-6C).

To further investigate the affect of PQBP1 on neurons, we focused on the VASE-associated isoforms of NCAM-140 encoded by *Ncam1*, which has been implicated in neurite outgrowth and development. Ncam1 contains five immunoglobulin-like (Ig) domains involved in homophilic binding of Ncam1 either in *cis* or *trans* that in turn controls cell-cell interactions and the induction of neurite outgrowth (Walsh and Doherty 1997). The AS of VASE determines the inclusion or exclusion of the fourth Ig domain and alters the homophilic binding of NCAM-140 (Lahrtz et al. 1997). The VASE – isoform of NCAM-140 stimulates — while VASE + inhibits — neurite outgrowth (Doherty et al. 1992; Saffell et al. 1994; Liu et al. 1993).

We found that the ratio of VASE –/VASE + expression corresponded to the level of functional PQBP1 and neurite outgrowth (Figure 3-4D; 3-5C ;3-7C,E,F). Moreover, restoration of the proper ratio of VASE –/VASE + 'rescued' the abnormal neurite outgrowth associated with the knockdown of PQBP1 (Figure 3-7F,G). These results identified NCAM-140 as a critical target connecting PQBP1-mediated AS to the cellular process of neurite projection and outgrowth. However, the ability of enforced VASE – /VASE + expression to rescue the neurite outgrowth defect caused by PQBP1 downregulation does not exclude the possibility that other PQBP1 targets contribute to this phenotype. For example, PQBP1 AS targets are also enriched in ARF-related proteins (Figure 3-6C). The members of the ARF family of small GTPases were

previously reported to affect dendritic branching and spine formation (Hernández-Deviez et al. 2002; Moore et al. 2007).

A possible role for PQBP1 in neurological disorders

Clinical genetics identified mutations in PQBP1 as the direct cause for Renpenning Syndrome, a representative XLMR disorder characterized by microcephaly, short stature, multiple body malformations, and mental retardation (Kalscheuer et al. 2003; Stevenson et al. 2005). We examined the function of two disease-associated PQBP1 variants—a truncated protein resulting from a ΔAG deletion, and a point mutation containing a Y65C mutation in the WW domain. We discovered that the disease-associated PQBP1 mutants had deficiencies in splicing-related functions. Our proteomics profiling indicates that the association between splicing-related factors and WT PQBP1 was lost or altered in the Y65C and ΔAG mutants, respectively (Table 3-1). These results indicate that the association of PQBP1 with splicing factors is important for its function. In line with their compromised association with the splicing machinery, neither ΔAG nor Y65C restored the AS of NCAM-140 VASE when PQBP1 was knocked down in mouse primary neurons (Figure 3-7C). Furthermore, both mutants were also unable to rescue the defects in neurite outgrowth upon depletion of PQBP1 (Figure 3-5E,F). Defects in dendritic development are consistently found in many other mental retardation disorders, including Fragile-X mental retardation and Rett Syndrome (Parrish et al. 2007). Our results suggest that disease-associated PQBP1 variants are deficient in maintaining the proper function of PQBP1 in AS. Moreover, aberrant splicing of specific AS targets may contribute to the disease symptoms in PQBP1-related mental retardation

disorders.

In summary, we propose that PQBP1, a protein related to neurological diseases, affects AS through association with the core splicing complex. Aberrant AS of specific targets resulting from misregulation of PQBP1 mutants may be one of the causes of PQBP1-related neurological disorders.

MATERIALS AND METHODS

Plasmids, cell culture, and siRNA transfection

Details of plasmid and stable cell-line constructions are included in Supplemental Materials and Methods. HeLa and 293T cell lines (ATCC) were grown in standard conditions. siRNAs were transfected with Hiperfect (Qiagen) at a final concentration of 40nM. siRNAs were siGENOME siRNAs (Thermo scientific) for PQBP1 and firefly luciferase. Transfections were optimized to reach 85% or more protein depletion 72 hours post-transfection, as determined by Western blotting.

Mouse cortical culture, plasmid transfection and lentivirus infection

Mouse embryonic cortical neurons were prepared and seeded as in Kim et al. 2010. For direct neuron transfection, MISSION® shRNA lentivirus constructs for PQBP1 (Sigma) and a non-targeting shRNA vector (Addgene Plasmid 1864) were transfected into neurons on the day of seeding with Lipofectamine2000 (Invitrogen). For lentiviral infection of neurons, 293T cells were transfected with shRNA constructs and packaging vectors. Lentivirus-containing supernatants were collected 48 hours post-transfection.

Neurons were then incubated with the supernatants for 6 hours on the day of seeding. Infection was optimized to reach 60% or more protein depletion 72 hours post-infection, as determined by Western blotting.

RNA extraction and RT-PCR

Cellular RNA was extracted by RNeasy Mini Kit (Qiagen). 1.5-4 μg RNAs were reverse-transcribed by Superscript First-Strand Synthesis System (Invitrogen). *Bcl-x* and NCAM-140 VASE splicing isoforms were analyzed by RT-PCR through primers flanking the alternatively spliced region. A linear range for cDNA input was determined by running a two fold dilution series of 28 cycle PCR reactions. Products were run on 5% Tris-Borate-EDTA gels (Biorad), stained with SYBR gold stain (Invitrogen) and quantified with Image J software (NIH). Triplicate samples in the linear range were run to determine isoform ratios. Real-time PCR primers were designed with the software Primer3 (SourceForge) and verified by electrophoresis to generate single, correctly sized products. PCR reactions were prepared with SYBR GREEN master mix (Applied Biosystems), run on a Mastercycler® ep realplex (Eppendorf) and quantified by the ΔΔC(t) method. Primer sets are listed in Supplemental Table S3-2.

Tandem affinity Immunopurification

FLAG-HA tagged PQBP1 and the Y65C and ΔAG mutants were purified from HeLa cell extracts by tandem affinity purification and visualized by silver stain as recently described (Adelmant et al. 2012). Specific details are provided in Supplemental Materials and Methods.

Sample preparation and mass spectrometry analysis

Sample preparation and LC-MS/MS analysis were done with slight modifications to a recently described method (Rozenblatt-Rosen et al. 2012). Specific details are provided in Supplemental Materials and Methods.

Western blotting

Cellular proteins were extracted with 1xRIPA buffer (IX PBS, 1% Triton X-100, 0.5% sodium deoxycholate, 0.1% SDS, protease inhibitors). 10-30 μg of proteins were loaded onto each lane of a NuPAGE 4-12% Bis-Tris gel. Proteins were transferred onto PVDF membranes and probed with the following primary antibodies diluted in 1X PBST with 5% non-fat milk: α-PQBP1 (Santa Cruz); α-FLAG (Sigma); α-HA (Roche); α-actin (Chemicon); α-SF3B1 (MBL).

CLIP

HeLa cells were cross-linked at 4000 mJ/cm^2 with a UV Stratelinker (Stratgene) and suspended in 1X RIPA buffer supplemented with protease inhibitors, 10mM DTT and 0.4U/μl RNAseOUT (Invitrogen). Lysates were treated briefly with RQ1 RNAse-free DNAse (Promega) and cleared by centrifugation. Dynal beads (Invitrogen) that were pre-coupled with 10μg of antibody or non-relevant IgG were incubated with the supernatants and rocked at 4°C for 4 hours. Beads were then washed 3 times with the lysis buffer plus 300mM NaCl and twice with 1X PNK buffer (50 mM Tris-Cl pH 7.4, 10 mM MgCl2, 0.5% NP-40). RNA-protein complexes were eluted and treated with

Protease K (Ambion) and RNAs were extracted by acid phenol/chloroform. Reverse transcription was performed with a 1:1 mix of oligo-dT and random hexamers and cDNAs were analyzed by real-time PCR.

Immunofluorescence

Neurons were fixed with 4% paraformaldehyde and 2% sucrose/PBS for 8 minutes. Cells were then washed 3 times with 1X PBS and treated with 1X GDB (*Gelatin Dilution Buffer*) buffer (0.3% triton, 0.1% gelatin) for 15 minutes. Cells were then incubated with antibodies, stained with DAPI and imaged on a Nikon Ti Inverted Fluorescence Microscope. For co-staining, we swapped the fluorophores on secondary antibodies to confirm there was no bleed-through in the fluorescence channels.

Annexin V staining

Neurons were incubated with FITC-conjugated Annexin V (BioLegend) diluted 1:50 in media at 37 °C for half an hour. Neurons were then directly imaged alive on a Nikon Ti Inverted Fluorescence microscope.

GO functional analysis

GO term enrichments for proteins associated with PQBP1 in HeLa cells were determined with DAVID Bioinformatics Database (Huang et al. 2009; Dennis et al. 2003). Functional GO enrichments of PQBP1 AS targets in neurons were determined with GoMiner™ software (Zeeberg et al. 2003).

ACKNOWLEDGMENTS

We thank Hsi-Wen Liao, Athar Malik, David Lipton and Seth Margolis for help with neuron culture and RNAi; Alan Mardinly for help with Sholl analysis; Jennifer Waters for help with imaging; Daniel Ducat, Natalie Farny and Yvonne Chen for thoughtful critiques. We are indebted to Michael Greenberg's lab for the generous sharing of mouse embryonic cortical neurons. Microscope imaging was done in the Nikon Imaging Center at Harvard Medical School. Sequencing was performed by the biopolymer facility at Harvard Medical School. This work was supported by grants GM36373 and GM057476 to P.A.S from the National Institutes of Health.

REFERENCES

Adelmant G, Calkins AS, Garg BK, Card JD, Askenazi M, Miron A, Sobhian B, Zhang Y, Nakatani Y, Silver PA, et al. 2012. DNA ends alter the molecular composition and localization of ku multicomponent complexes. *Mol Cell Proteomics* **11:** 411-421.

Black DL. 2003. Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* **72:** 291–336.

Akgul C, Moulding DA, Edwards SW. 2004. Alternative splicing of Bcl-2-related genes: functional consequences and potential therapeutic applications. *Cell Mol Life Sci* **61:** 2189–2199.

Castle JC, Zhang C, Shah JK, Kulkarni AV, Kalsotra A, Cooper TA, Johnson JM. 2008. Expression of 24,426 human alternative splicing events and predicted cis regulation in 48 tissues and cell lines. *Nat Genet* **40:** 1416–1425.

Chakarova CF, Hims MM, Bolz H, Abu-Safieh L, Patel RJ, Papaioannou MG, Inglehearn CF, Keen TJ, Willis C, Moore AT, et al. 2002. Mutations in HPRP3, a third member of pre-mRNA splicing factor genes, implicated in autosomal dominant retinitis pigmentosa. *Hum Mol Genet* **11:** 87-92.

Charizanis K. Lee KY, Batra R, Goodwin M, Zhang C, Yuan Y, Shiue L, Cline M, Scotti MM, Xia G, et al. 2012. Muscleblind-like 2-mediated alternative splicing in the developing brain and dysregulation in myotonic dystrophy. *Neuron* **75:** 437-450.

Dennis G Jr., Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA. 2003. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* **4:** P3.

Doherty P, Moolenaar CE, Ashton SV, Michalides RJ, Walsh FS. 1992. The VASE exon downregulates the neurite growth-promoting activity of NCAM 140. *Nature* **356:** 791–793.

Dredge BK, Polydorides AD, Darnell RB. 2001. The splice of life: alternative splicing and neurological disease. *Nat Rev Neurosci* **2:** 43–50.

Edgar R, Domrachev M, Lash AE. 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30:** 207-210.

Gehman LT, Stoilov P, Maguire J, Damianov A, Lin CH, Shiue L, Ares M Jr., Mody I, Black DL. 2011. The splicing regulator Rbfox1 (A2BP1) controls neuronal excitation in the mammalian brain. *Nat Genet* **43:** 706-711.

Gozani O, Potashkin J, Reed R. 1998. A potential role for U2AF-SAP 155 interactions in recruiting U2 snRNP to the branch site. *Mol Cell Biol* **18:** 4752-4760.

Grabowski PJ, Black DL. 2001. Alternative RNA splicing in the nervous system. *Prog Neurobiol* **65:** 289–308.

Hernández-Deviez DJ, Casanova JE, Wilson JM. 2002. Regulation of dendritic development by the ARF exchange factor ARNO. *Nat Neurosci* **5:** 623-624.

Huang DW, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nature Protoc* **4:** 44-57.

Kalscheuer VM, Freude K, Musante L, Jensen LR, Yntema HG, Gécz J, Sefiani A, Hoffmann K, Moser B, Haas S, et al. 2003. Mutations in the polyglutamine binding protein 1 gene cause X-linked mental retardation. *Nat Genet* **35:** 313–315.

Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptewicz M, Barbara-Haley K, Kuersten S, et al. 2010. Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465:** 182-187.

Komuro A, Saeki M, Kato S. 1999. Association of two nuclear proteins, Npw38 and NpwBP, via the interaction between the WW domain and a novel prolinerich motif containing glycine and arginine. *J Biol Chem* **274:** 36513–36519.

Kunde SA, Musante L, Grimme A, Fischer U, Muller E, Wanker EE, Kalscheuer VM. 2011. The X-chromosome-linked intellectual disability protein PQBP1 is a component of neuronal RNA granules and regulates the appearance of stress granules. *Hum Mol Genet* **20:** 4916–4931.

Licatalosi DD, Darnell RB. 2006. Splicing regulation in neurologic disease. *Neuron* **52:**

93–101.

Liu L, Haines S, Shew R, Akeson RA. 1993. Axon growth is enhanced by NCAM lacking the VASE exon when expressed in either the growth substrate or the growing axon. *J Neurosci Res* **35:** 327–345.

Lunde BM, Moore C, Varani G. 2007. RNA-binding proteins: modular design for efficient function. *Nat Rev Mol Cell Biol* **8:** 479–490.

Makarova OV, Makarov EM, Urlaub H, Will CL, Gentzel M, Wilm M, Lührmann R. 2004. A subset of human 35S U5 proteins, including Prp19, function prior to catalytic step 1 of splicing. *EMBO J* **23:** 2381–2391.

Massiello A, Roesser JR, Chalfant CE. 2006. SAP155 binds to ceramide-responsive RNA cis-element 1 and regulates the alternative 5' splice site selection of Bcl-x pre-mRNA. *FASEB J* **20:** 1680–1682.

McKie AB, McHale JC, Keen TJ, Tarttelin EE, Goliath R, van Lith-Verhoeven JJ, Greenberg J, Ramesar RS, Hoyng CB, Cremers FP, et al. 2001. Mutations in the pre-mRNA splicing factor gene PRPC8 in autosomal dominant retinitis pigmentosa (RP13). *Hum Mol Genet* **10:** 1555-1562.

Moore MJ, Wang Q, Kennedy CJ, Silver PA. 2010. An alternative splicing network links cell-cycle control to apoptosis. *Cell* **142:** 625–636.

Moore CD, Thacker EE, Larimore J, Gaston D, Underwood A, Kearns B, Patterson SI, Jackson T, Chapleau C, Pozzo-Miller L, et al. 2007. The neuronal Arf GAP centaurin alpha1 modulates dendritic differentiation. *J Cell Sci* **120:** 2683-2693.

Musante L, Kunde S-A, Sulistio TO, Fischer U, Grimme A, Frints SGM, Schwartz CE, Martínez F, Romano C, Ropers H-H, et al. 2010. Common pathological mutations in PQBP1 induce nonsense-mediated mRNA decay and enhance exclusion of the mutant exon. *Hum Mutat* **31:** 90–98.

Okazawa H, Rich T, Chang A, Lin X, Waragai M, Kajikawa M, Enokido Y, Komuro A, Kato S, Shibata M, et al. 2002. Interaction between mutant ataxin-1 and PQBP-1 affects transcription and cell death. *Neuron* **34:** 701–713.

Okazawa H, Sudol M, Rich T. 2001. PQBP-1 (Np/PQ): a polyglutamine tract-binding and nuclear inclusion-forming protein. *Brain Res Bull* **56:** 273-280.

Parrish JZ, Emoto K, Kim MD, Jan YN. 2007. Mechanisms that regulate establishment, maintenance, and remodeling of dendritic fields. *Annu Rev Neurosci* **30:** 399–423.

Qi Y, Hoshino M, Wada Y-I, Marubuchi S, Yoshimura N, Kanazawa I, Shinomiya K-I, Okazawa H. 2005. PQBP-1 is expressed predominantly in the central nervous system during development. *Eur J Neurosci* **22:** 1277–1286.

Rozenblatt-Rosen O, Deo RC, Padi M, Adelmant G, Calderwood MA, Rolland T, Grace M, Dricot A, Askenazi M, Tavares M, et al. 2012. Interpreting cancer genomes using systematic host network perturbations by tumor virus proteins. *Nature* **487:** 491-495.

Saffell JL, Walsh FS, Doherty P. 1994. Expression of NCAM containing VASE in neurons can account for a developmental loss in their neurite outgrowth response to NCAM in a cellular substratum. *J Cell Biol* **125:** 427–436.

Sholl DA. 1953. Dendrit.ic organization in the neurons of the visual and motor cortices of the cat. *J Anat* **87:** 387-406.

Small SJ, Akeson R. 1990. Expression of the unique NCAM VASE exon is independently regulated in distinct tissues during development. *J Cell Biol* **111:** 2089–2096.

Spector DL, Lamond, AI. 2011. Nuclear speckles. *Cold Spring Harb Perspect Biol* **3:** pii: a000646.

Stevenson RE, Bennett CW, Abidi F, Kleefstra T, Porteous M, Simensen RJ, Lubs HA, Hamel BCJ, Schwartz CE. 2005. Renpenning syndrome comes into focus. *Am J Med Genet* **134:** 415–421.

Tapia VE, Nicolaescu E, McDonald CB, Musi V, Oka T, Inayoshi Y, Satteson AC, Mazack V, Humbert J, Gaffney CJ, et al. 2010. Y65C missense mutation in the WW domain of the Golabi-Ito-Hall syndrome protein PQBP1 affects its binding activity and

deregulates pre-mRNA splicing. *J Biol Chem* **285:** 19391–19401.

Ule J, Jensen KB, Ruggiu M, Mele A, Ule A, Darnell RB. 2003. CLIP identifies Nova-regulated RNA networks in the brain. *Science* **302:** 1212–1215.

Ule J, Ule A, Spencer J, Williams A, Hu J-S, Cline M, Wang H, Clark T, Fraser C, Ruggiu M, et al. 2005. Nova regulates brain-specific splicing to shape the synapse. *Nat Genet* **37:** 844–852.

Varani G, Nagai K. 1998. RNA recognition by RNP proteins during RNA processing. *Annu Rev Biophys Biomol Struct* **27:** 407–445.

Vithana EN, Abu-Safieh L, Allen MJ, Carey A, Papaioannou M, Chakarova C, Al-Maghtheh M, Ebenezer N.D, Willis C, Moore AT, et al. 2001. A human homolog of yeast pre-mRNA splicing gene, PRP31, underlies autosomal dominant retinitis pigmentosa on chromosome 19q13.4 (RP11). *Mol. Cell* **8:** 375-381.

Wahl MC, Will CL, Lührmann R. 2009. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136:** 701–718.

Walsh FS, Doherty P. 1997. Neural cell adhesion molecules of the immunoglobulin superfamily: role in axon growth and guidance. *Annu Rev Cell Dev Biol* **13:** 425–456.

Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB. 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature* **456:** 470–476.

Wang G-S, Cooper TA. 2007. Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat Rev Genet*. **8:** 749–761.

Waragai M, Lammers CH, Takeuchi S, Imafuku I, Udagawa Y, Kanazawa I, Kawabata M, Mouradian MM, Okazawa H. 1999. PQBP-1, a novel polyglutamine tractbinding protein, inhibits transcription activation by Brn-2 and affects cell survival. *Hum Mol Genet*. **8:** 977–987.

Waragai M, Junn E, Kajikawa M, Takeuchi S, Kanazawa I, Shibata M, Mouradian MM,

Okazawa H. 2000. PQBP-1/Npw38, a nuclear protein binding to the polyglutamine tract, interacts with U5-15kD/dim1p via the carboxylterminal domain. *Biochem. Biophys. Res. Commun*. **273:** 592–595.

Zeeberg BR, Feng W, Wang G, Fojo AT, Sunshine M, Narasimhan S, Kane DW, Reinhold WC, Lababidi S, Bussey KJ, et al. 2003. GoMiner: a resource for biological interpretation of genomic and proteomic data. *Genome Biol*. **4:** R28.

CHAPTER 4: A JUNCTION-BASED METHOD FOR DIFFERENTIAL ANALYSIS OF GLOBAL ALTERNATIVE SPLICING FROM RNA-SEQ DATA

ABSTRACT

Alternative splicing (AS) is widely recognized as a major source of biological diversity and regulation in eukaryotes. >90% of mammalian genes undergo AS and present tissue- and development stage-specific AS profiles. These characteristics make it necessary to study AS at a global level. High-throughput transcriptome sequencing (RNA-seq) is a powerful method to examine AS at a genome-wide scale. However, the diversity of AS patterns presents computational challenges to quantify AS from RNA-seq data. Currently, several methods are applied in AS quantification and differential AS analysis, such as Cuffdiff, MISO and DEXSeq. These methods use RNA-seq reads mapped to a full splicing transcript, or key AS regions of a transcript, or exons of a gene. Most of these methods rely on a priori knowledge of the genomic annotations of exons and cannot be applied in examining tissue-specific novel AS events. These methods also lack sensitivity necessary to detect changes in AS events that involve short alternatively spliced exons. Here, we present a new method to analyze global AS profiles under different conditions based on RNA-seq reads mapped to splice junctions only. Our method does not depend on current genomic annotation and presents a thorough and sensitive comparison of AS patterns between conditions. When applied to quantify global AS changes upon depletion of PQBP1 in mouse cortical neurons, our method identified many AS targets of PQBP1 that were later experimentally validated. Interestingly, some experimentally verified targets, such as *Ncam1* that has an alternatively spliced exon of 30bp could only be recognized by our method but not other software. These results showed the specificity and sensitivity of our junction-based AS quantification method.

INTRODUCTION

RNA-seq as a powerful tool to investigate AS on a global scale

Alternative splicing (AS) is a major source of biological diversity in eukaryotes. Greater than 90% of mammalian genes encode pre-mRNAs that undergo AS and generate more than one transcript (Wang et al. 2008; Pan et al. 2008) using one or more of the following seven basic AS patterns: cassette exon inclusion/exclusion, mutually exclusive exon usage, alternative 5' splice site usage, alternative 3' splice site usage, alternative promoter site usage, alternative poly-A site usage and intron retention (Black 2003; Figure 1-2). In addition to adding great diversity to the transcriptome in multicellular eukaryotes, AS also plays critical regulatory roles in biological processes, and in many cases acts as the crucial determinant for cell fate and identity. The regulation of AS is carried out by a wide range of RNA-binding proteins (RBPs) that associate with *cis*-elements embedded in precursor messenger RNA (pre-mRNA) sequences (Black 2003). AS regulation is further incorporated into cellular signaling pathways and responds dynamically to environmental cues (Moore et al. 2008; Wahl et al. 2009).

One of the major challenges in studying AS is to systematically quantify and compare global AS changes under different conditions. Indeed, the disturbance of specific AS regulators can lead to aberrant global AS changes that are closely associated with human diseases (Ule et al. 2005; Gehman et al. 2011; Charizanis et al. 2012). Such changes in global AS patterns can only be recognized by comparing genome-wide AS profiles between the healthy and abnormal states. High-throughput transcriptome sequencing (RNA-seq) is a powerful means of investigating AS at the genome-wide scale. However, the diversity of AS patterns and complex hierarchical AS regulation

bring forward computational challenges to quantify global AS changes from RNA-seq data.

Current bioinformatics tools to quantify AS

RNA-seq on a biological sample generates a massive amount of short reads (length varying from 50bp to 200bp) that represent the whole transcriptome and can be used to quantify the expression levels of different mRNA transcripts from each gene. The common practice for RNA-seq–based differential analysis is to first define a quantification unit, such as a gene or a specific transcript from a gene. The read counts of each quantification unit can then be collected and compared between conditions using statistical tools (Pepke et al. 2009). Various software packages have been developed to analyze differential gene expression patterns from RNA-seq data between conditions by comparing the collective count of reads that map to any mRNA isoforms from a given gene (Garber et al. 2011). Analyzing the AS changes of a gene's mRNAs, however, is a more complicated problem. The major issue lies in the proper definition of the quantification unit that represents the AS patterns of a gene's pre-mRNAs. So far, only a few bioinformatics tools have been developed for differential AS analysis with various definitions of the quantification unit. In general, these quantification units can be summarized into two categories.

The first category uses a full splicing isoform or the key AS region of a splicing isoform as the quantification unit for AS (Figure 4-1A; Mortazavi et al. 2008; Wang et al. 2008). Tools in this category include Cuffdiff (Trapnell et al. 2010; Trapnell et al. 2013) and MISO (Katz Y et al. 2010). This definition of quantification unit depends on a priori

Figure 4-1: Bioinformatics models for AS quantification.

(A) A schematic of AS quantification from RNA-seq data is shown for a hypothetical alternatively spliced pre-mRNA. Green rectangles represent exons and black lines between exons are introns. For models that use the whole splicing isoform or key AS region of that isoform as the quantification unit of AS, black short lines mark reads that are mapped to the full isoform. (B) For models that use exons as the quantification unit of AS, blue short lines mark reads that are mapped to each exon of the pre-mRNA for exon usage quantification between conditions. (C) For our method that use splice junction as the quantification unit, red short lines mark reads that are mapped to each splice junction of the mRNA.

(D) A schematic of a defined AS event and quantification model of AS in our method. The middle exon is the alternatively spliced exon. 5' IS is the start of a splicing junction and 3' ES the end of the junction. The AS event shown here includes two sub-AS-junctions that share the same 5' IS. Red short lines mark RNA-seq reads mapped to sub-AS-junction-1 and purple short lines mark reads mapped to sub-AS-junction-2. The distribution of reads mapped to the two sub-AS-junctions was used for differential analysis of AS in our analysis.

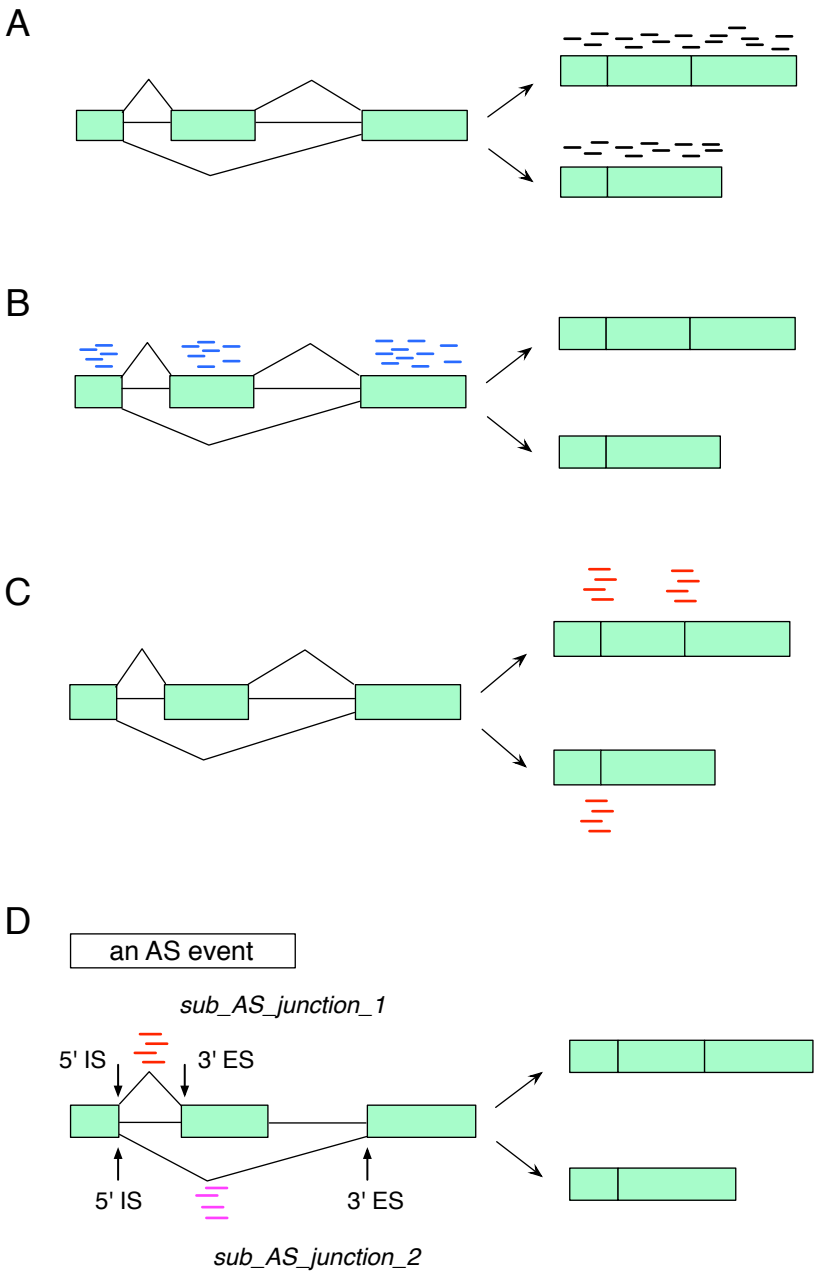

Figure 4-1 (Continued).

knowledge of all the mRNA isoforms/AS patterns of a gene. There are two ways to get priori knowledge of these mRNA isoforms and AS patterns across the genome. One is to use current genomic annotation (Wang et al. 2008; Katz et al. 2010) to extract information about all known splicing patterns or mRNA isoforms of a gene. However, this approach has major problems: novel AS events will not be considered. The current annotations of organism transcriptomes are far from complete, as high-throughput studies constantly identify novel AS events in different samples (Mortazavi et al. 2008). Novel and especially tissue-specific AS events are in many cases crucial, since various studies have shown that functional AS decisions are highly dynamic and are usually made in a tissue- and developmental stage-specific manner (Wang et al. 2008; Mortazavi et al. 2008). The other way is to reconstruct the transcriptome of a specific biological sample from the RNA-seq data so that information about mRNA splicing isoforms of all genes can be collected. Although this approach is able to include novel splicing isoforms, reconstruction of the transcriptome from massive amount of short reads is itself a computationally challenging problem (Garber et al. 2011). Many genes in higher organisms produce more than 10 or even thousands of isoforms and have a combination of multiple basic AS patterns, especially in complex tissues such as the brain (Nilsen and Graveley 2010). As a result, even with the most powerful computational tools, it is difficult to decipher every transcript of a given gene simply from the sequencing data. The other shortcoming of this approach is the lack of the sensitivity necessary for analyzing changes brought by small alternatively spliced exons/exonic regions. In general, the length of a transcript (several thousand bp) is much larger than that of an alternatively spliced exon (10 to a few hundred bp), so a full-length transcript can attract

far more reads than a short alternatively spliced exon. As a result, especially in long transcripts, changes in read count of a transcript contributed by an AS event that involves small alternatively spliced exons are relatively hard to detect.

The second category of AS quantification tools define the quantification unit based on exon usage (Figure 4-1B). Usage, defined as the number of reads mapped to each exon of the gene, can be calculated and compared among different conditions. Tools of this category include DEXSeq (Anders and Huber 2010). The detection of differential exon usage under specific conditions indicates changes in the exclusion or inclusion of exons, and thus AS patterns of that gene (Anders and Huber 2010). This exon-based approach also relies on a priori knowledge of an annotated genome and has the same problem as discussed in the first category. In addition, this method is not sensitive enough to detect differential usage of short exons (some exons can be as short as 20bp), especially if the gene under investigation has many other long exons. The reason is that a short exon cannot attract many read counts in the first place, especially when the exon length is shorter than the read length. The changes in read count from AS of that exon will be small compared to the read counts from other exons.

In fact, the most straightforward means of measuring AS changes is to examine the number of reads that are directly mapped to different AS junctions (Figure 4-1C). However, this approach has not been applied to quantify global AS changes from RNA-seq data.

Biological variability and differential analysis

Variability among biological replicates is crucial for generating valid statistics in differential analysis (Hansen et al. 2011). Due to costs and technical difficulties, only a small number of replicates are performed in RNA-seq experiments, usually not enough to estimate biological variability among samples through standard approaches (Garber et al. 2011). As a result, the majority of current computational methods for global AS differential analysis exclude biological variability in their statistics and can only compare one sample out of several replicates in each condition in a pairwise manner (Anders et al. 2012; Wang et al. 2008; Katz et al. 2010). Interestingly, methods to infer biological variability from a small number of replicates have been developed for sequencing technologies (Robinson and Smyth 2007; Robinson and Smyth 2008), and have started to be implemented for differential analysis using RNA-seq, such as software packages Edge R, DESeq and DEXSeq (Robinson et al. 2010; Anders and Huber 2010; Anders et al. 2012). These methods have been tested in practice for differential gene expression analysis and have proven to be effective in modeling and incorporating biological variability for generating reliable statistics (Becker et al. 2011; Hah et al. 2011; Ross-Innes et al. 2012).

We hereby present a computational method that use only reads directly mapped over splice junctions to compare differential AS profiles from RNA-seq data. Applying this method, we identified 457 AS events that experience significant changes upon PQBP1 depletion in mouse embryonic cortical neurons. Subsequent experimental validation on the computationally predicted AS targets of PQBP1 demonstrates the high specificity and sensitivity of our method.

RESULTS

Description of the junction-based method

We considered RNA-seq reads that are directly mapped over splice junctions as the quantification unit for splicing changes (Figure 4-1C). In fact, before RNA-seq, exon junction microarrays that bear probes straddling selected splice junctions have been used for AS quantification (Johnson et al. 2003). Computational methods comparing AS profiles using exon junction arrays gave reliable results on differential analysis (Ule et al. 2005). For each splice junction we define the 5' initiation site (5' IS) of the junction as the final nucleotide position of the upstream exon and the 3' end site (3' ES) as the first nucleotide position of the downstream exon (Figure 4-1D). An AS event is thus defined as a set of splice junctions that share the same 5' IS or the same 3' ES, with each junction in the set referred to as a sub-AS-junction (Figure 4-1D). This definition of AS covers all 6 primary AS patterns except for intron retention, which was not considered in our analysis (Figure 4-2; Black 2003). The splicing pattern of an AS event under a certain condition is represented by the distribution of read counts mapped respectively to each sub-AS-junction within the AS event—i.e., the 'usage' of each sub-AS-junction under that condition.

We take a two-step approach to retrieve read counts on splice junctions. RNA-seq reads represent the transcriptome, i.e. the processed transcripts where introns are already removed. Thus the reads that are from exonic regions can be mapped directly back to the genome, while reads that straddle splice junctions cannot. Our first step is to identify all the splice junctions and the junction sequences. Reads are mapped to the genome using Tophat, an alignment program that maps short RNA-seq reads to the genome and

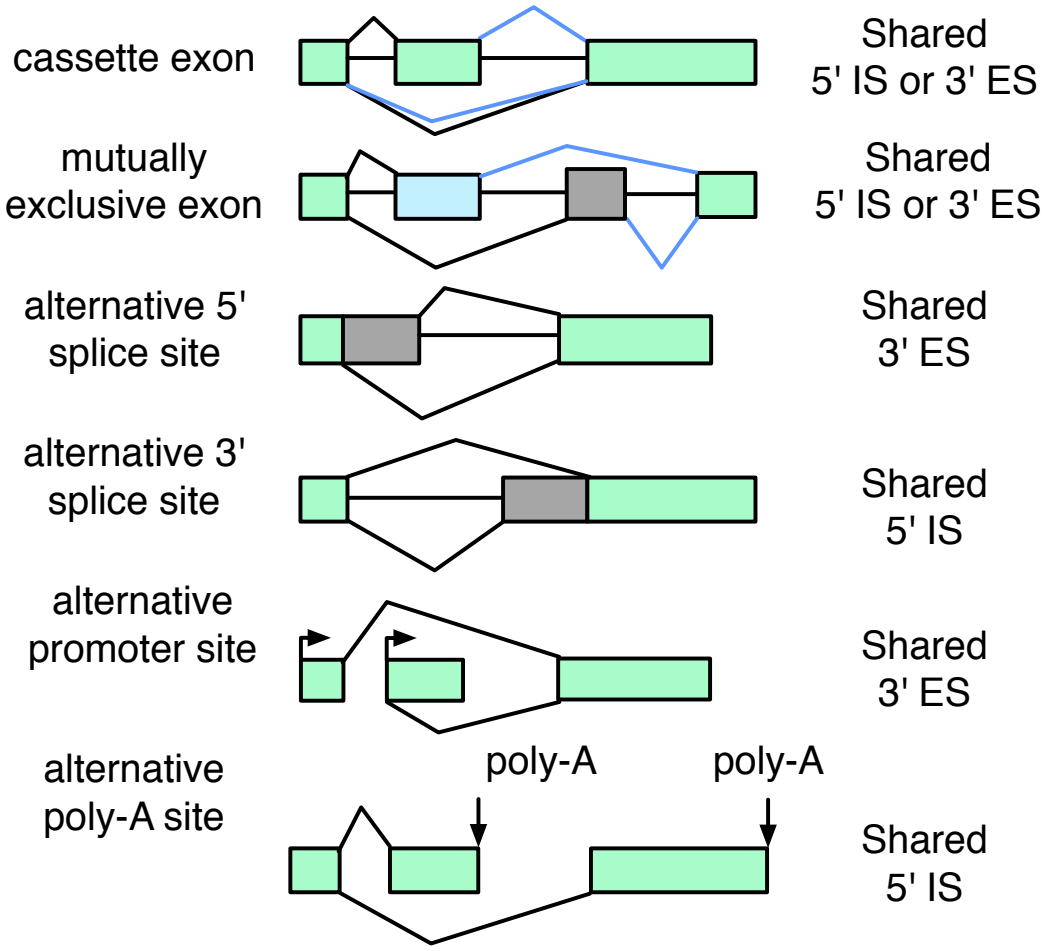

Figure 4-2: Definition of AS events in our method covers basic AS patterns.

identifies exon-exon junctions without a priori knowledge of the exon annotation (Trapnell et al. 2009). Tophat recognizes splice junctions by first directly mapping fully mappable reads to the genome to define "exon islands". After that, reads that straddle splice junctions and are thus not directly mappable in the first step are systematically

segmented and each segment is mapped separately to different exon islands, so as to discover separate exons or exonic regions that are straddled by these reads, i.e. the splice junction (Trapell et al. 2009). In this way, Tophat provides a thorough scan for not only known splice junctions, but also novel junctions in the specific sample. After junction identification, we generate a junction sequence database by extracting exonic sequences upstream and downstream of each junction. The second step is to collect read counts on each identified splice junction. We combine the mouse genome sequence and the junction sequence database together and use Bowtie, a short read alignment program (Langmead et al. 2009) to map all the reads to the combined sequence database, this time without read segmentation. In this way, reads that are from exonic regions can be directly mapped to the genome and reads that span a splice junction can be directly mapped to the profiled junction sequences, and read counts for each splice junction are achieved.

To analyze the differential 'usage' of each sub-AS-junction within an AS event between conditions, we apply a modified version of the method DEXSeq (Anders et al. 2012). DEXSeq was originally developed to analyze differential exon usage within a gene under different conditions. Dependent on prior knowledge of genome annotation, DEXSeq utilizes biological variability among replicates to generate robust statistical tests directly from raw read counts (Robinson and Smyth 2007; Anders and Huber 2010; Anders et al. 2012). We adapted this method to analyze the differential 'usage' of each sub-AS-junction in an AS event between different conditions. Since our statistical tools are adapted from DEXSeq, we incorporated the variability among biological replicates in our analysis.

Application of the junction-based method

As an application of this method, we compared global AS profiles between control and PQBP1-knockdown (PQBP1 KD) mouse cortical neuron samples as described in Chapter 3. RNA samples from three non-targeting shRNA treated replicates and two PQBP1 KD replicates were extracted and sequenced. Each sample under each condition received 55-120 million paired-end 50bp reads.

Reads were mapped to the mouse genome (GRCm38/mm10) using Tophat and a total of 141,493 splice junctions were identified from the mouse embryonic cortical neuron samples (Figure 4-3). A junction sequence database was then established by extracting exonic sequences upstream and downstream of each junction from the UCSC genome browser (Karolchik et al. 2004); the exact junction length varied from 46bp to 98bp, depending on the Tophat report for the span of the junction. We used Bowtie2 to map all the reads to the combined sequence database of the mouse genome and the splice junction sequence database (Figure 4-3). For each sample, ~80% of reads were mapped uniquely and for multi-hit reads we picked the top-scored mapping site, so that ~98% of reads mapped to the combined genome and junction database.

We prepared Perl scripts to profile AS events and the corresponding sub-AS-junctions from the splice junction database identified by Tophat, as defined previously. The read counts on each sub-AS-junction were extracted respectively with Samtools (Li et al. 2009). Perl scripts were written to adapt the R package of DEXSeq (Anders et al. 2012; Gentleman et al. 2004) to analyze the differential usage of sub-AS-junctions within an AS event, comparing control and PQBP1 KD samples. To make the analysis more stringent, only sub-AS-junctions with one or more mapped reads from all five control and

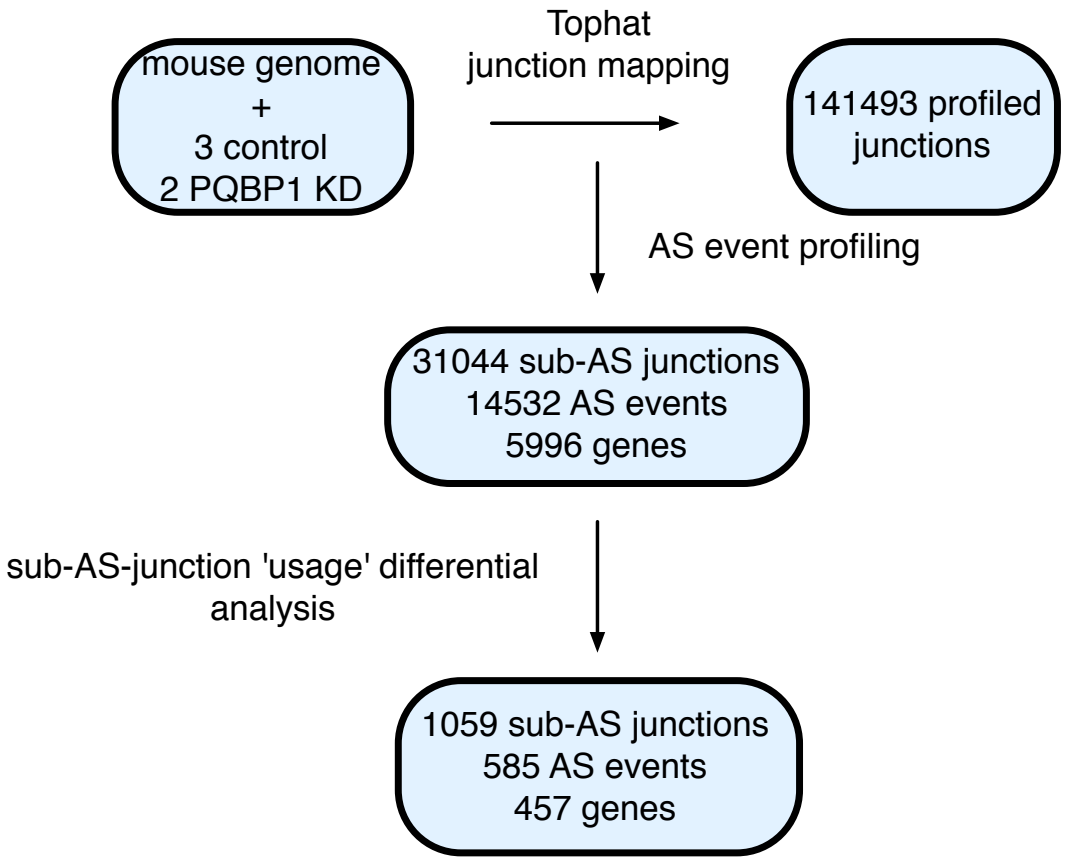

Figure 4-3: Flow chart of the computational analysis to identify AS targets of PQBP1 from RNA-seq data.

PQBP1 KD samples were considered. A sub-AS junction that received less than ten reads adding up from all five samples was excluded from statistical tests, as suggested by DEXSeq (Anders et al. 2012). From 141,493 junctions identified by Tophat, our profiling resulted in 14,532 AS events involving 31,044 sub-AS-junctions, distributed among 5996 genes (Figure 4-3). From there, 585 AS events were identified to exhibit significant

changes in usage of sub-AS-junctions upon PQBP1 KD with a p-value cutoff at 0.05, covering 457 genes (Supplemental Table S4-1; Figure 4-3).

Validation of the junction-based method

To validate our method, we randomly picked ten computationally predicted PQBP1 AS targets and two predicted non-targets, and analyzed the splicing of each candidate by RT-PCR (Figure 3-6B). All ten predicted targets showed significant changes in AS and the predicted non-targets did not (Figure 3-6B), indicating the specificity and sensitivity of our junction-based method to analyze differential AS between different conditions.

DISCUSSION

In this work, we described a new junction-based computational method to efficiently analyze differential global AS from RNA-seq data. We used RNA-seq reads that are directly mapped over splice junctions as the quantification unit and defined the distribution of read counts among sub-AS-junctions within an AS event as the AS pattern of that event. We also applied statistical methods that incorporate variability among biological replicates to justly analyze the differential usage of sub-AS-junctions of an AS event under different conditions. We applied this method in evaluating genome-wide AS changes in neurons upon PQBP1 depletion and validated our results experimentally. Broadly, our method provides a thorough and reliable way to quantify global AS changes between different conditions and can be applied to many aspects of AS studies such as identifying genome-wide AS targets of a splicing regulator.

Read distribution among splicing junctions as a rational quantification unit for AS

The number of reads that map over a splice junction is a direct and quantitative reflection of how much that splice site is selected during the process of pre-mRNA AS. Although it is difficult to reconstruct every transcript that a gene generates from RNA-seq data, there has been significant progress in designing algorithms that identify splice junctions of the transcriptome (Trapnell et al. 2009; Au et al. 2010; Wang et al. 2010). Tophat is among the most well developed and widely used algorithms within this category (Trapnell et al. 2009). In our method, Tophat provides a relatively complete profiling of all the splice junctions as the basis for achieving accurate read counts for each junction.

Our method defines an AS event as a set of sub-AS-junctions that share the same 5' IS or 3' ES and describe AS pattern of this event as the distribution of reads among sub-AS-junctions within that AS event. This definition conveys similar ideas as Ule et al. 2005 with exon junction microarrays in which each AS event is examined between different samples by comparing the hybridization intensity of probes representing each sub-AS-junction. In Ule et al., 49 computationally predicted AS events that experienced the most significant changes upon Nova1 knockout were tested by RT-PCR and all 49 were validated (Ule et al. 2005). In our case, all 10 randomly picked predicted PQBP1 AS targets and 2 non-targets were also verified by RT-PCR. These results indicate the high specificity and sensitivity of our model for AS quantification using read distribution among sub-AS-junctions.

Moreover, our method adapts DEXSeq to model biological variability and incorporate it into statistical testing for significant changes in sub-AS-junction usages within AS events between conditions. In this way we avoided discrepancies that may be brought up by pairwise comparisons among samples under different conditions.

A thorough and sensitive detection of global AS changes

Our method can detect changes in AS events that are tissue-specific and have not been annotated before. For example, we computationally predicted and experimentally validated an important AS target of PQBP1 using our method—the splicing of the VASE (Variably Alternatively Spliced Exon) exon in *Ncam1* between exon 7 and 8 (Figure 3-7A; 3-6B). The VASE-excluding isoform of *Ncam1* (VASE –) promotes neurite outgrowth while the VASE-including isoform (VASE +) inhibits neurite outgrowth (Doherty et al. 1992). Thus, *Ncam1* is an important component among PQBP1's AS targets that functionally relates the role of PQBP1 in AS regulation to the phenotypic activity of neurite outgrowth (See chapter 3). Although reported before in the literature, the VASE exon is not included in the current mouse genome annotation (Figure 4-4). As a result, computational methods such as MISO and DEXSeq that depend on priori knowledge of annotated genomes will not be able to identify *Ncam1* as PQBP1's AS target. On the other hand, methods that perform differential AS analysis by reconstructing the transcriptome from RNA-seq such as Cuffdiff (Trapnell et al. 2010; Trapnell et al. 2013) could not identify the AS of VASE in *Ncam1* either (data not shown). One possible reason could be the difficulty of correctly reconstructing all the AS transcripts of *Ncam1* from RNA-seq reads.

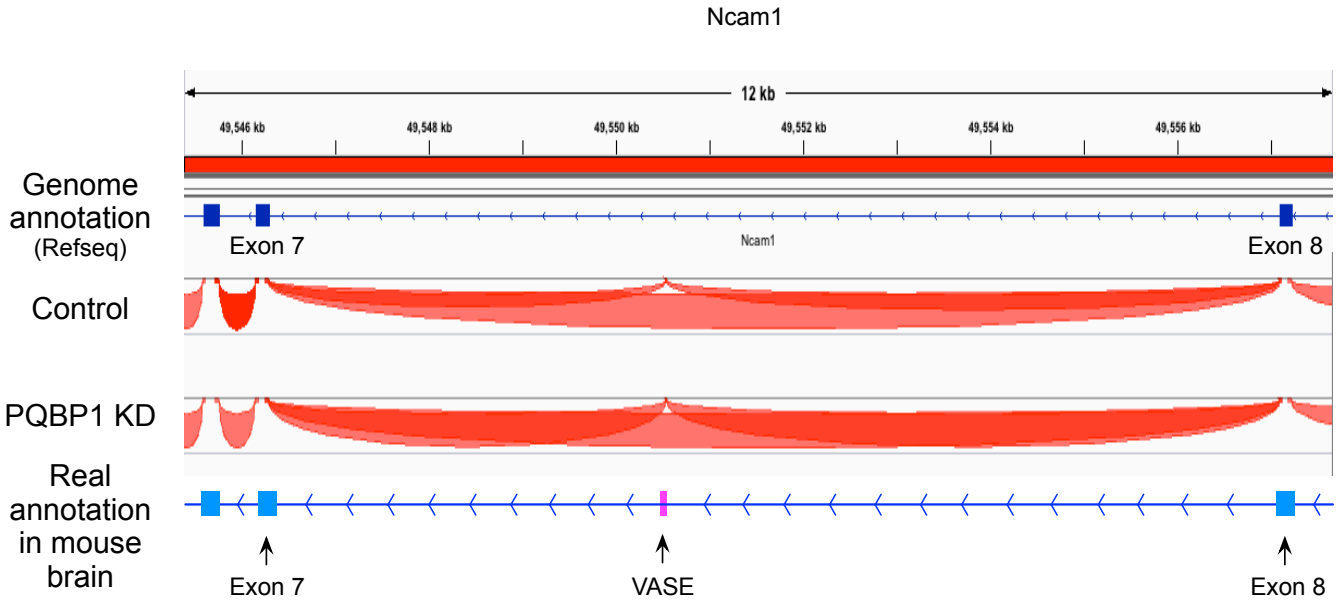

Figure 4-4: RNA-seq reads mapped to junctions between exon7 and VASE, and VASE and exon 8 in mouse gene *Ncam1* from control and PQBP1 KD samples.

Genomic structure of mouse *Ncam1* in shown. Blue rectangles are exons and lines between exons are introns, with arrowheads on introns showing the direction of transcription. Genomic structure on top is from the Refseq database, while the one on the bottom is the *Ncam1* genomic structure in mouse brain inferred from RNA-seq data. VASE exon (pink rectangle) is not included in Refseq but is discovered from RNAseq data. The middle two lanes show the reads count from control and PQBP1 KD samples on splicing junctions. Each red arc between exons represents reads mapped to the corresponding splice junction. The height and thickness of an arc show the relative depth of read counts mapped to that junction. In PQBP1 KD sample, the arcs connecting exon 7 and VASE and VASE and exon 8 are significantly increased in comparison to the control sample, depicting an increase in the generation of VASE + isoform. This figure is generated by the Integrative Genomics Viewer (Thorvaldsdóttir et al. 2013; Robinson et al. 2011).

Our method also shows better sensitivity in detecting AS changes than other algorithms. The lengths of splicing junctions identified from Tophat are approximately similar to each other. As a result, when examining the relative usage of each sub-AS-junction within an AS event, a change in the read count mapped to a sub-AS-junction is

quick to detect. Take *Ncam1* as an example again. The VASE exon is only 30bp long while the length of an *Ncam1* transcript is around 6kb. If using the transcript as the quantification unit, the read count from the inclusion or exclusion of VASE can hardly cause noticeable changes in the total amount of reads received by an *Ncam1* transcript. This may be another reason that Cuffdiff could not identify *Ncam1* as an AS target of PQBP1. As for exon-based AS differential analysis methods such as the original DEXSeq, even adding the annotation for VASE could not improve the prediction result of DEXSeq on *Ncam1*. It is again possible that the short VASE exon compared to other exons has a small number of reads mapped to it in the first place, and that the change in read count caused by the AS of VASE is too weak to be detected.

In contrast, when using reads mapped to the splicing junctions as the quantification unit, a significant difference in VASE splicing is observed immediately between the control and PQBP1 KD samples (Figure 4-4). These facts again illustrate the sensitivity of our junction-based method in performing differential AS analysis between different conditions.

MATERIALS AND METHODS

RNA-seq sample preparation

We used the PrepX mRNA IntegenX kit (Illumina) to prepare library constructions from total RNA. The detailed protocol can be tracked through the link provided below:

http://www.genome.med.harvard.edu/documents/illumina/IntegenX_Apollo324_mRNA_Seq_Protocol_10012012.pdf

Samples were then run on Illumina HiSeq 2000. The reported sequencing data has been deposited in NCBI's Gene Expression Omnibus (Edgar et al. 2002) and is accessible through GEO series accession number GSE44402.

ACKNOWLEDGEMENT

Computational analysis on RNA-seq data was analyzed on the Orchestra shared research cluster provided by the Research Information Technology Group at Harvard Medical School. This work is supported by grants GM36373 and GM057476 from the National Institutes of Health.

REFERENCES

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol* **11:** R106.

Anders S, Reyes A, Huber W. 2012. Detecting differential usage of exons from RNA-seq data. *Genome Res* **22:** 2008-2017.

Au KF, Jiang H, Lin L, Xing Y, Wong WH. 2010. Detection of splice junctions from paired-end RNA-seq data by SpliceMap. *Nucleic Acids Res* **38:** 4570-4578.

Black DL. 2003. Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* **72:** 291–336.

Becker C, Hagmann J, Müller J, Koenig D, Stegle O, Borgwardt K, Weigel D. 2011. Spontaneous epigenetic variation in the Arabidopsis thaliana methylome. *Nature* **480:** 245-249.

Charizanis K, Lee KY, Batra R, Goodwin M, Zhang C, Yuan Y, Shiue L, Cline M, Scotti MM, Xia G, et al. 2012. Muscleblind-like 2-mediated alternative splicing in the developing brain and dysregulation in myotonic dystrophy. *Neuron* **75:** 437-450.

Edgar R, Domrachev M, Lash AE. 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30:** 207-210.

Garber M, Grabherr MG, Guttman M, Trapnell C. 2011. Computational methods for transcriptome annotation and quantification using RNA-seq. *Nat Methods* **8:** 469-477.

Gehman LT, Stoilov P, Maguire J, Damianov A, Lin CH, Shiue L, Ares M Jr., Mody I, Black DL. 2011. The splicing regulator Rbfox1 (A2BP1) controls neuronal excitation in the mammalian brain. *Nat Genet* **43:** 706-711.

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, et al. 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5:** R80.

Hah N, Danko CG, Core L, Waterfall JJ, Siepel A, Lis JT, Kraus WL. 2011. A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. *Cell* **145:** 622-634.

Hansen KD, Wu Z, Irizarry RA, Leek JT. 2011. Sequencing technology does not eliminate biological variability. *Nat Biotechnol* **29:** 572-573.

Johnson JM, Castle J, Garrett-Engele P, Kan Z, Loerch PM, Armour CD, Santos R, Schadt EE, Stoughton R, Shoemaker DD. 2003. Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* **302:** 2141-2144.

Karolchik D, Hinrichs AS, Furey TS, Roskin, KM, Sugnet CW, Haussler D, Kent WJ. 2004. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res* **32:** D493-496.

Katz Y, Wang ET, Airoldi EM, Burge CB. 2010. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat Methods* **7:** 1009-1015.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10:** R25.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* **25:** 2078-2079.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5:** 621–628.

Moore MJ, Silver PA. 2008. Global analysis of mRNA splicing. *RNA* **14:** 197–203.

Nilsen TW, Graveley BR. 2010. Expansion of the eukaryotic proteome by alternative splicing. *Nature* **463:** 457-463.

Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* **40:** 1413–1415.

Pepke S, Wold B, Mortazavi A. 2009. Computation for ChIP-seq and RNA-seq studies. *Nat Methods* **6:** S22-32.

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotechnol* **29:** 24-26.

Robinson MD, Smyth GK. 2007. Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics* **23:** 2881-2887.

Robinson MD, Smyth GK. 2008. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics* **9:** 321-332.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edger: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26:** 139-140.

Ross-Innes CS, Stark R, Teschendorff AE, Holmes KA, Ali HR, Dunning MJ, Brown GD, Gojis O, Ellis IO, Green AR, et al. 2012. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481:** 389-393.

Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14:** 178-192.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-seq. *Bioinformatics* **25:** 1105-1111.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28:** 511-515.

Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol* **31:** 46-53.

Ule J, Ule A, Spencer J, Williams A, Hu J-S, Cline M, Wang H, Clark T, Fraser C, Ruggiu M, et al. 2005. Nova regulates brain-specific splicing to shape the synapse. *Nat Genet* **37:** 844–852.

Wahl MC, Will CL, Lührmann R. 2009. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136:** 701–718.

Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB. 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature* **456:** 470–476.

Wang K, Singh D, Zeng Z, Coleman SJ, Huang Y, Savich GL, He X, Mieczkowski P, Grimm SA, Perou CM, et al. 2010. MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic Acids Res* **38:** e178.

CHAPTER 5: A SYNTHETIC ALTERNATIVE SPLICING NETWORK TO CONFER MEMORY OF EXTRACELLULAR STIMULI IN MAMMALIAN CELLS

ABSTRACT

Alternative splicing (AS) of mRNA is a major source of biological diversity and regulation in eukaryotes. Despite its rapidity and flexibility as a regulatory mechanism, AS remains under-utilized in synthetic biology. The eukaryotic AS regulatory network contains multiple standard network motifs including positive and negative feedback loops. In many cases, AS functions as a switch modulating critical biological functions, such as cell fate and identity determination during development in multicellular eukaryotes. AS allows cells to respond to environmental cues at the post-transcriptional level, often more rapidly than through transcriptional regulation alone. Here, we demonstrate a synthetic splicing-based circuit in mammalian cells—based on the autoregulatory feedback loop from the splicing regulation of Sex-Lethal (*Sxl*), the *Drosophila* sex determination master gene—that confers memory of extracellular stimuli. We show that, as in *Drosophila*, Sxl regulates the splicing of its own mRNA in mammalian cells. Furthermore, through quantitative modeling and experimental data, we show that this positive feedback loop can serve as a crucial element in synthetic devices to rapidly respond to and retain memory of transient exposure to stimuli in mammalian cells. Our splicing-based memory device explores the potential of AS regulatory networks in synthetic circuit design and will contribute to the study of long-term effects of transient stimuli on cells as well as our knowledge of post-transcriptional regulation in biology.

INTRODUCTION

In biological systems, it is common that transient stimuli lead to long-term effects, especially in cellular decision-making and cell fate determination. As an example, the NOTCH signaling pathway regulates the maintenance of neural progenitor cells as well as their differentiation into neurons and glia (Gaiano and Fishell 2002). A brief inactivation of NOTCH is adequate to shift a series of downstream molecular events and cause progenitor cells to commit to permanent differentiation (Nelson et al. 2007). The importance of transient stimuli in long-term cell dynamics is also prominent in human diseases. For instance, hypoxia is associated with many pathological conditions including cancer, and acute or chronic hypoxia can bring different biological effects in tumor cells (Bristow and Hill 2008).

Cellular decision-making dependent on transient stimuli arises from bistability, the phenomenon of the existence of two stable states of a network that are switchable in the presence of appropriate stimuli (Strogatz 1994). In certain cases, bistability shows irreversibility, as in the case of cells permanently committing to a specific fate or state upon receiving a transient signal (Thomas and Kaufman 2001). In other words, cells keep the 'memory' of this signal, change the gene expression profile accordingly, and self-perpetuate this expression profile. Irreversibility in bistability is usually achieved by strong positive feedback loops in the regulatory network (Xiong and Ferrell 2003).

With traditional approaches, it is difficult to identify cells that have been influenced by past transient stimuli. Synthetic biology on the other hand, provides tools to design and engineer circuits that confer cellular 'memory' of a past transient signal into a readout that is experimentally detectable (Ajo-Franklin et al. 2007). The protracted

cellular response to a transient signal in the past can then be studied by tracking the cells that carry markers of memory. Currently, synthetic bistable circuits are mostly based on feedback loops adapted from transcriptional regulatory networks (Gardner et al. 2000; Atkinson et al. 2003; Kramer et al. 2004; Vilaboa et al. 2005; Ajo-Franklin et al. 2007; Ingolia and Murray 2007; Burrill and Silver 2011; Burrill et al. 2012). In addition, most of these circuits are implemented in prokaryotes, as the establishment of bistability and memory in eukaryotes, especially in mammalian cells, has proved to be more difficult (Ingolia and Murray 2007; Burrill et al. 2012). The few transcription-based synthetic memory circuits built in eukaryotes show heterogeneous performance on memory retention in the population (Burrill and Silver 2011; Burrill et al. 2012). It has been proposed that this phenomenon is due to the kinetic features of transcriptional regulation, especially the time scale of transcription activation compared to that of other cellular activities such as cell division (Burrill and Silver 2011).

Alternative splicing (AS) is a crucial component in biological regulatory networks. AS regulatory network is part of post-transcriptional regulation and allows cells to respond to environmental cues more rapidly than through mechanisms at the transcriptional level alone. In addition, AS regulatory network is composed of multiple standard network motifs including negative and positive feedback loops, and often functions as the major determinant of cell fate.

Perhaps the most striking example of AS determining cell fate comes from the splicing of Sex-lethal (*Sxl*) encoding the master regulator for *Drosophila* sex determination (Figure 5-1). The primary determinant of fly sex is the ratio between the number of X chromosomes and the number of sets of autosomes (X:A) in the cell. In

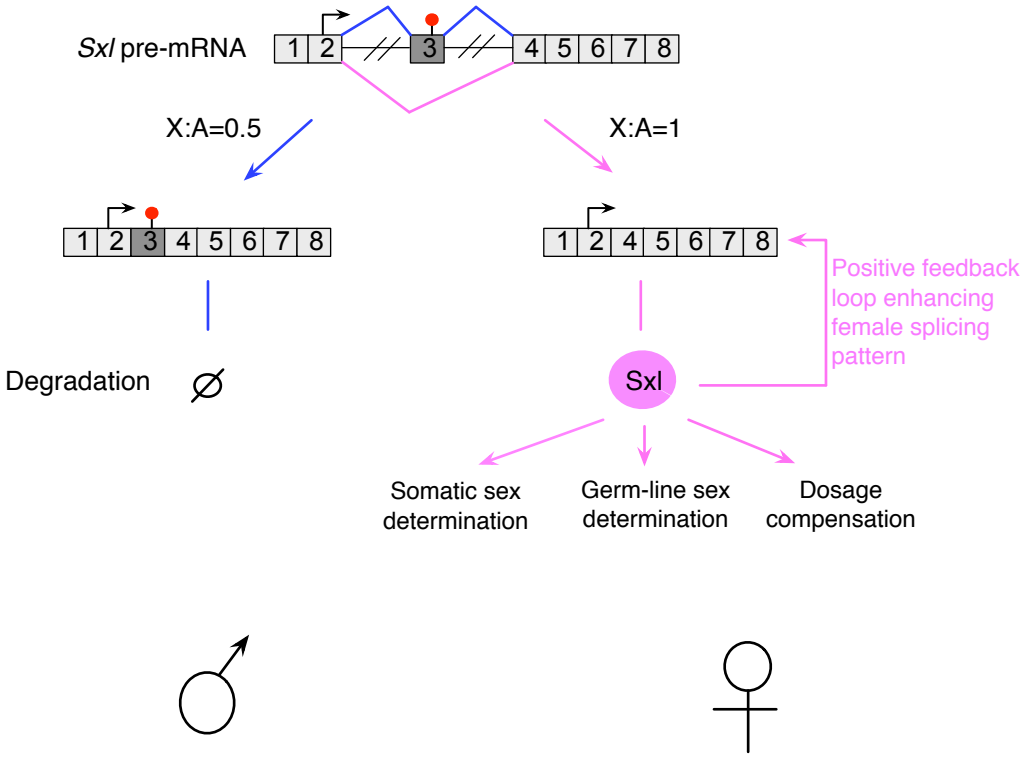

Figure 5-1: Alternative splicing regulation of *Sxl* and sex determination in Drosophila.

A schematic of the AS of Drosophila *Sxl* is shown. Gray rectangles are exons and black lines between exons are introns; the dark rectangle marks the alternatively spliced exon 3; arrows show translation start sites; stop signs show premature termination codons. In male flies where the ratio of number of X chromosome to sets of autosomes (X:A) is 0.5, *Sxl* follows the default splicing pattern to include exon 3 that contains a stop codon, leading to transcript degradation. In female flies where X:A is 1, *Sxl* is spliced to skip exon 3 and generates mature Sxl proteins. Sxl is itself a splicing regulator and enhances the female-specific pattern to keep a steady production of Sxl proteins. Sxl affects the functions of a cascade of downstream genes and thereby regulates female-specific phenotype determination processes such as somatic sex determination, germ line sex determination and dosage compensation.

male flies where the X:A ratio is 0.5, the splicing of *Sxl* includes exon 3 that contains a premature termination codon. Thus the male-specific *Sxl* transcripts are degraded through nonsense-mediated mRNA decay (NMD) and do not produce Sxl protein. In contrast, in

female flies where the X:A ratio is 1, an alteration in the splicing of *Sxl* is initiated early in embryogenesis to skip exon 3 and leads to the expression of full-length Sxl protein. This protein is itself a splicing factor that auto-regulates the splicing of *Sxl* to maintain the female-specific isoform, resulting in a steady production of Sxl protein. Besides its own pre-mRNA, Sxl also regulates the splicing of pre-mRNAs of transformer (tra) and male-specific lethal-2 (msl-2), and modulate the functions of a cascade of downstream regulators for male- or female-specific phenotypes including somatic sex determination, germ-line sex determination and dosage compensation (Figure 5-1; Baker 1989; Penalva and Sánchez, 2003). Through this mechanism, throughout a fly's life cells maintain the memory of sex information via the positive feedback loop of Sxl regulating the AS of its own pre-mRNA. Considering the homogenous establishment of sex memory in every fly cell throughout development and the fast cell cycle (8 minutes on average for embryonic cells) in fly cells, it is possible that the AS regulation of Sxl possesses unique kinetic features that allow fast and uniform action.

With a diverse collection of regulatory network motifs and rapid response to signals, AS is currently underutilized in synthetic circuit design. Here, we present a synthetic mammalian memory circuit that utilizes the AS regulation of *Sxl* to confer memory of transient doxycycline (dox) exposure. We show that the self-regulation of splicing of Sxl is preserved in mammalian cells, as in *Drosophila*, and that the auto-regulatory loop of Sxl splicing regulation can serve as a functional module of a synthetic circuit to rapidly confer memory of extracellular signals. We further performed quantitative stochastic modeling of the memory circuit for rational device optimization and integration into mammalian cells. Collectively, our study explores the potential of AS

regulatory network applications in synthetic devices and contributes to the understanding of post-transcriptional regulation in biology.

RESULTS

Splicing of Drosophila Sxl is preserved in mammalian cells

We found that splicing of *Drosophila Sxl* is preserved in mammalian cells (Figure 5-2A,B,C,D). To test if the expression of Sxl is harmful to mammalian cells, we first transfected U2OS cells with GFP-tagged mature Sxl. No noticeable negative effect on U2OS cells was observed. In addition, Sxl localizes in the nucleus of U2OS cells as in flies (Figure 5-2A).

The default Sxl splicing pattern in male flies includes exon 3 with a stop codon, which leads to transcript degradation. Only the expression of functional Sxl protein itself can shift the splicing of *Sxl* to the female-specific isoform that excludes exon 3. To determine whether the AS regulation of *Sxl* is preserved in mammalian cells, we co-transfected U2OS cells with a *Sxl* AS plasmid—a plasmid that contains CMV-driven *Sxl* fused to *GFP* with alternatively spliced exon 3 and adjacent introns (Figure 5-2B)—and an mCherry-tagged mature Sxl. As a control, we co-transfected the *Sxl* AS plasmid with a plasmid that expresses mCherry by itself (Figure 5-2B). Cells were then analyzed by flow cytometry (Figure 5-2B,C) and Western blotting (Figure 5-3D) for GFP-tagged Sxl expression. A significant increase in GFP-tagged Sxl was detected in the co-transfection of the *Sxl* AS plasmid with the mCherry-tagged Sxl in comparison to with mCherry alone (Figure 5-2B,C,D). These results illustrate that the splicing pattern of *Sxl* is the same in

Figure 5-2: Splicing of *Sxl* in mammalian cells.

(A) Fluorescence microscopy examination of U2OS cells transfected with GFP-tagged Sxl, in comparison to transfection with just GFP.

(B) A schematic of a *Sxl* AS plasmid is shown. Grey rectangles are exons; dark rectangle marks the alternatively spliced exon 3 and its adjacent intronic sequences are represented by black lines. Arrow shows translation start site and stop sign shows the premature termination codon. *Sxl* in this plasmid is fused to GFP, represented by a green rectangle. The inclusion or exclusion of exon 3 in *Sxl* splicing determines whether functional GFP-tagged Sxl can be generated. U2OS cells were co-transfected with the *Sxl* AS plasmid and mCherry-tagged mature Sxl, or the *Sxl* AS plasmid and mCherry. Cells were analyzed by flow cytometry. For each chart, X-axis is the intensity of the GFP fluorescence, and Y-axis is the intensity of the mCherry fluorescence. Percentage of cells falling into each quadrant (Q1, Q2, Q3, Q4) is listed.

(C) Quantification of the GFP fluorescence intensity of U2OS cells in Q2 (cells co-expressing GFP- and mCherry-tagged Sxl) as in (B).

(D) Western blot of U2OS cells that were co-transfected with the *Sxl* AS plasmid and mCherry-tagged mature Sxl, or the *Sxl* AS plasmid and mCherry, probed with GFP antibody. Size markers are listed on the left.

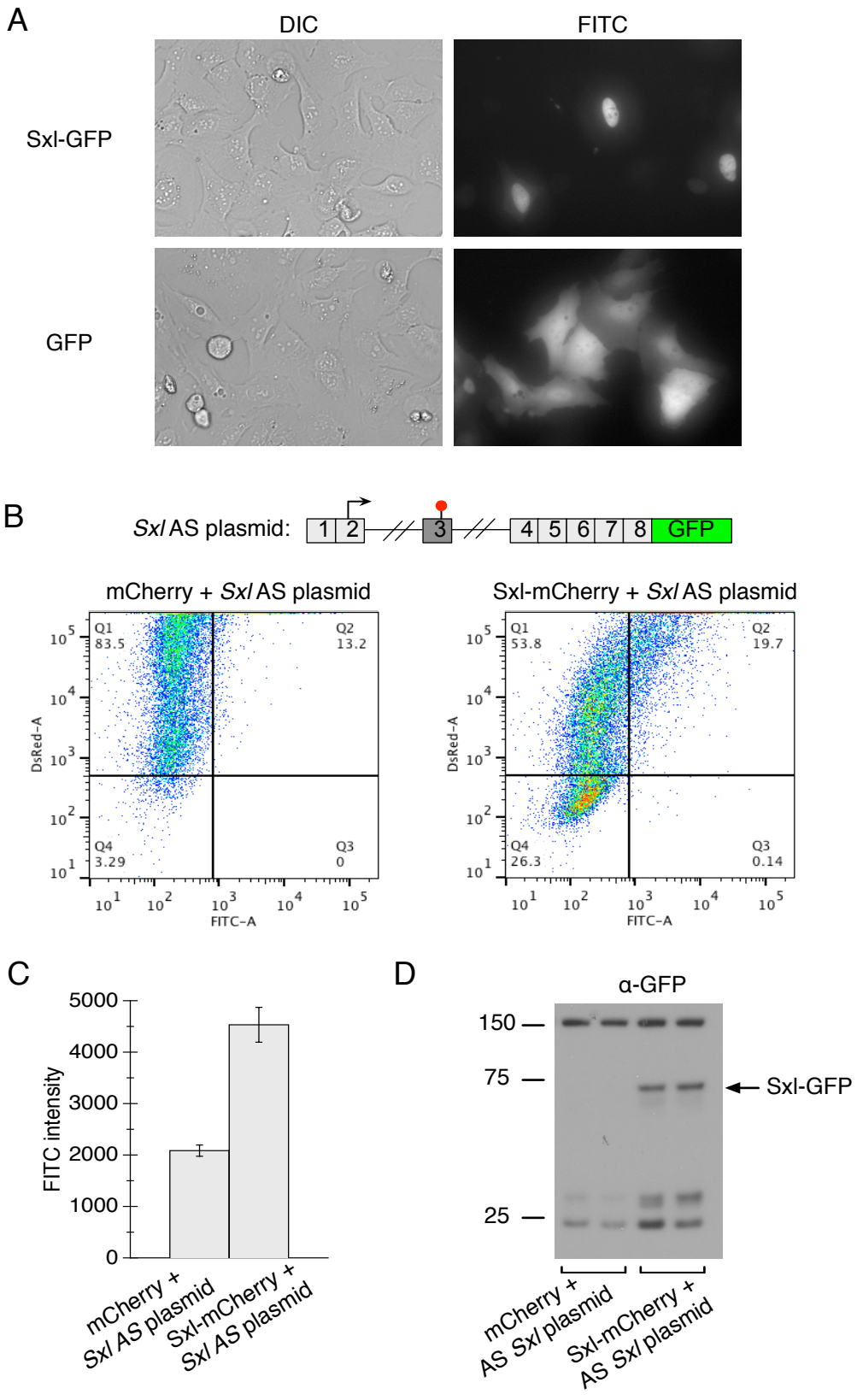

Figure 5-2 (Continued).

mammalian cells as in *Drosophila* and that the positive feedback loop of Sxl protein regulating its own pre-mRNA's splicing is preserved in mammalian cells.

Design of a splicing-based memory circuit

We designed a splicing-based memory circuit as shown in Figure 5-3A. For the trigger, the mature female-specific *Sxl* was cloned in C-terminal fusion with *mCherry*, driven by a dox-sensitive promoter T-REx. The T-REx promoter is repressed by the binding of Tet repressor that is previously integrated and continuously expressed in the specific U2OS cell lines we use (Blacklow Laboratory). When dox is present Tet repressor is removed from the promoter allowing expression of mCherry-tagged Sxl (Yao et al. 1998). For the reporter, the alternatively spliced exon 3 of *Sxl*, together with the adjacent introns and other constitutive exons (exon 1, 2, 4, 5, 6, 7 and 8) were cloned in C-terminal fusion with *GFP*, driven by a CMV promoter (Figure 5-3A). Without dox, the trigger does not produce any functional mCherry-tagged Sxl protein (Figure 5-3B). The *Sxl* transcripts generated from the CMV promoter of the reporter thus follow the splicing pattern in male flies that includes exon 3 and the premature termination codon. This leads to transcript degradation and no production of GFP-tagged Sxl proteins (Figure 5-3B). Upon dox exposure, the mCherry-tagged Sxl protein generated from the trigger will alter the splicing of *Sxl* transcripts from the reporter to exclude exon 3 and produce functional GFP-tagged Sxl proteins (Figure 5-3C). Sxl will in turn auto-regulate the splicing of its own pre-mRNA from the reporter and enhance the altered splicing pattern for continuous production of GFP-tagged Sxl, even after the trigger is turned off (Figure 5-3C). The GFP

Figure 5-3: Design of the splicing-based memory circuit and predicted responses.
(A) A schematic of the design of the *Sxl* splicing-based memory circuit. Grey rectangles are exons and dark rectangles marks the alternatively spliced exon 3 in *Sxl*. Black lines on either side of exon 3 represent the adjacent intronic sequences. Arrows show translation start site and stop signs show premature termination codons. The trigger is under the dox-sensitive promoter T-Rex. The reporter is driven by a CMV promoter.
(B) Device response without dox exposure. With no dox, the trigger does not express functional mCherry-tagged Sxl and the reporter follows the default male splicing pattern. No GFP-tagged Sxl protein is generated and no fluorescence signal is detected.
(C) Device response upon dox exposure. With dox, the trigger is activated and expresses mCherry-tagged mature Sxl protein, which in turn shifts the splicing of the reporter to exclude exon 3. GFP-tagged mature Sxl protein is generated from the reporter and self-enhances the reporter *Sxl* transcript to follow the female-specific splicing pattern.

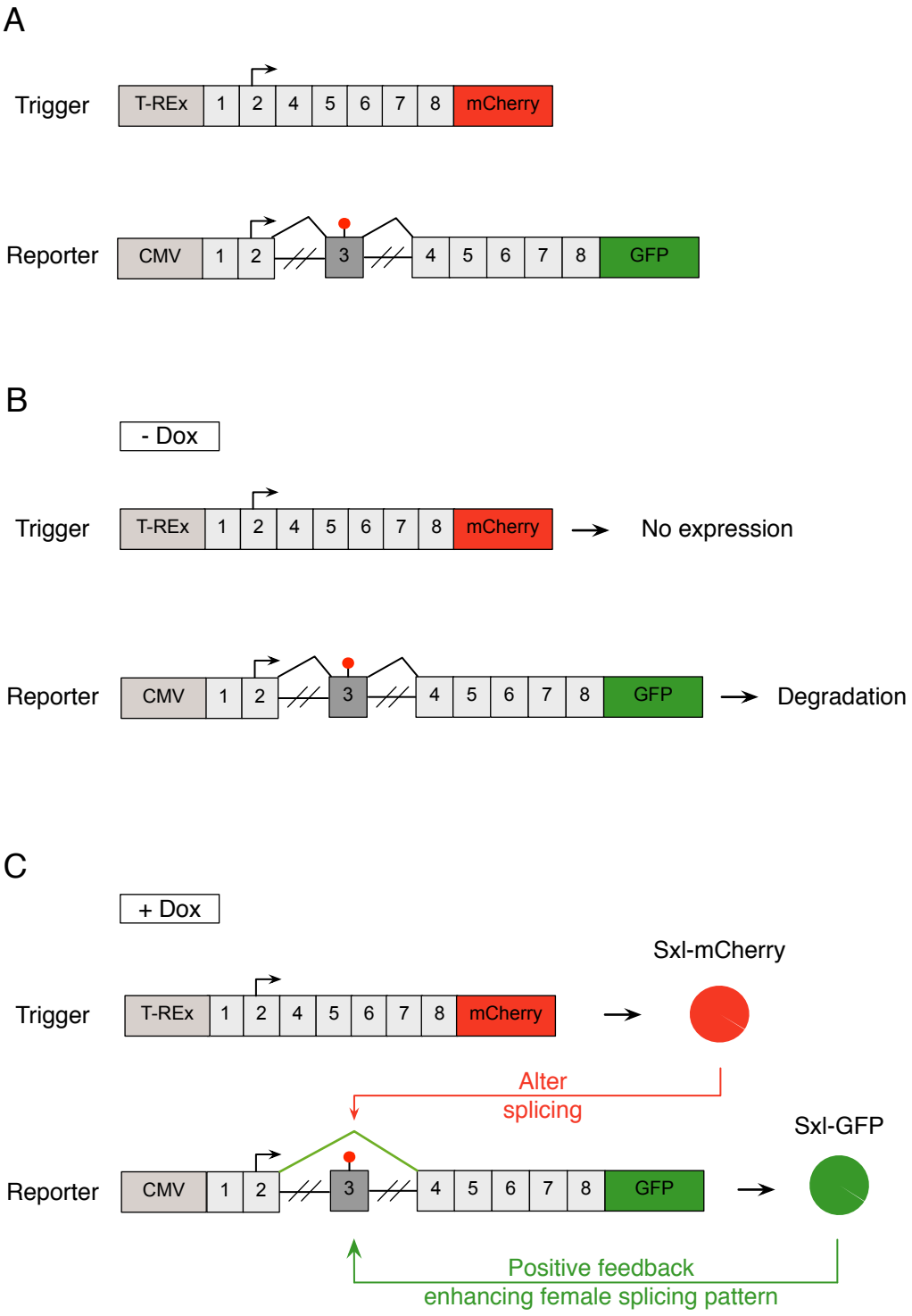

Figure 5-3 (Continued).

fluorescence signal here serves as a marker for memory of a transient dox exposure in the past.

Integration of the splicing-based memory circuit in U2OS cells

Stable integration of the memory circuit into mammalian cells is necessary for evaluation of the device's ability to maintain memory over time. By disassembling a circuit into several plasmids and randomly integrating them separately into the cells, stable cell lines are generated containing a varying number of copies of each circuit part in different positions of the genome. We first made U2OS cell lines that stably carry the proposed memory device by random integration of the trigger and reporter constructs on separate plasmids. The trigger plasmid expresses puromycin resistance and the reporter plasmid expresses hygromycin B resistance (Figure 5-4A). U2OS cells were co-transfected with the trigger and reporter plasmids and were dual-selected by puromycin and hygromycin B for cell clones that stably integrated both the trigger and the reporter of the device as a whole. Different cell clones were picked after the dual selection and each was examined for the device response.

The resulting stable U2OS cell line showed prompt activation of the trigger to dox exposure and rapid response from the reporter to the trigger (Figure 5-4B,C). U2OS cells were exposed to dox for 19 hours and examined for the fluorescence signal from the trigger and reporter by microscope. We observed that the trigger expressed high level of mCherry-tagged Sxl upon dox induction and the reporter generated GFP-tagged Sxl due to the activation of the trigger (Figure 5-4B). To validate that the reporter responds to the trigger by AS of *Sxl*, we designed RT-PCR primers flanking the alternatively spliced

Figure 5-4: Analysis of U2OS cells that stably integrated the memory circuit through random integration.

(A) Trigger and reporter constructs used for generating stable cell lines that carry the device. Arrowheads show RT-PCR primers used to analyze the splicing of the reporter Sxl transcript. (B) Trigger response (by DsRed for mCherry) and reporter activities (by FITC for GFP) are detected 19 hours post dox induction (19 hpi) in U2OS cells by fluorescence microscope. (C) RT-PCR analysis of Sxl transcripts generated from the reporter before and after dox induction. A significant shift from the isoform that contains exon 3 (bands in the right panel) to isoform that excludes exon 3 (bands in the left panel) is observed. Exon 3 is around 170 bp long so a band shift of 170 bp is expected.

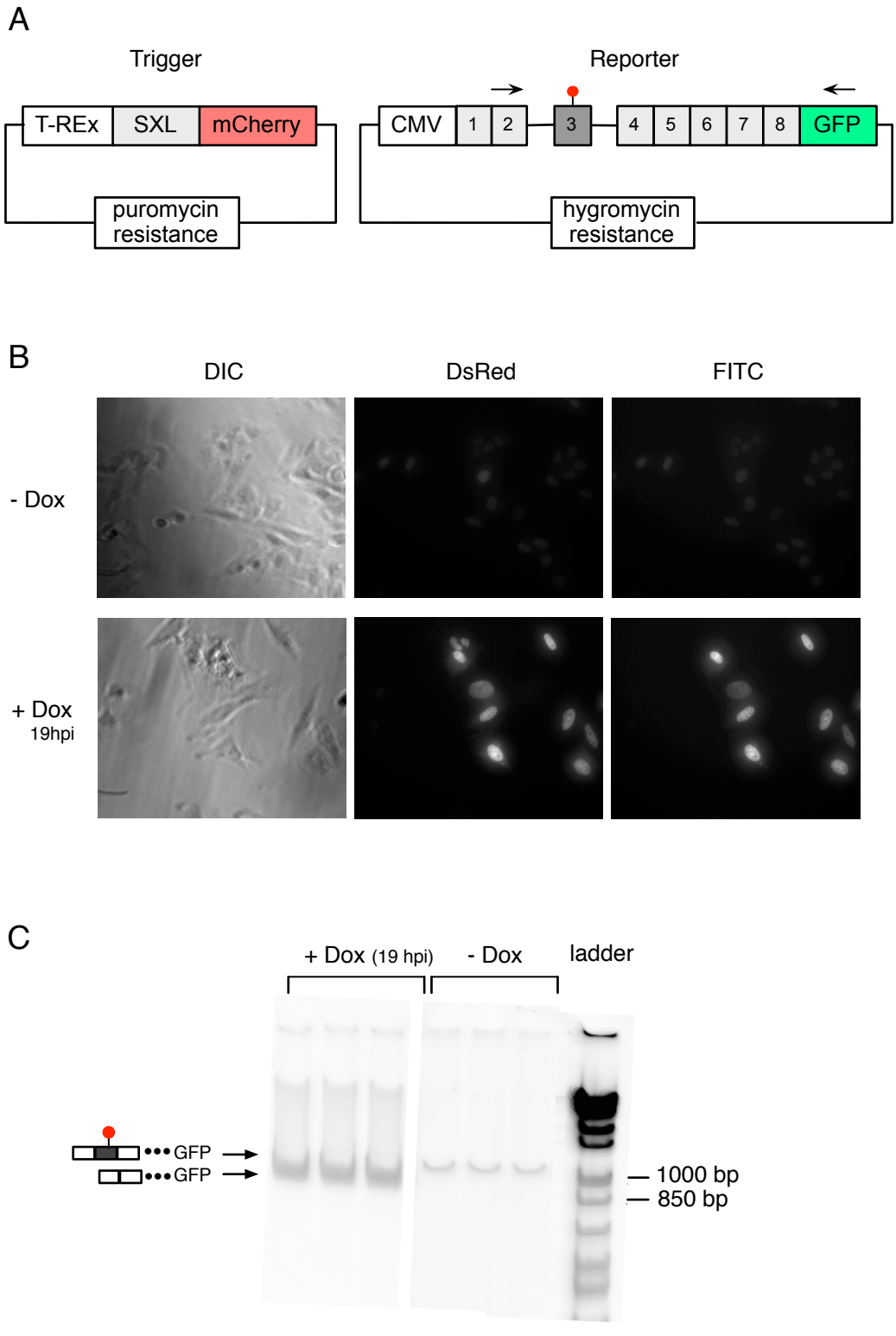

Figure 5-4 (Continued).

exon 3 (spanning to *GFP* for exclusive reporter *Sxl* transcripts; see Figure 5-4A). A significant shift in the splicing of reporter transcripts was observed from the exon 3-inclusive male-specific isoform to the exon 3-exclusive female-specific isoform upon the exposure to dox (Figure 5-4C). In the future, this stable cell line will be tracked for many days after removal of dox to observe memory of dox exposure.

Rational optimization of device design and genomic integration in U2OS cells

To optimize the long-term maintenance of memory, it is critical to investigate the dynamics of the splicing-based memory circuit to optimize the balance between the kinetics of trigger activation and the positive feedback loop of the reporter. Random integration of the device does not allow control of copy number and ratio of trigger/reporter constructs, or the integration location in the genome, making it difficult to evaluate and optimize the design of the device. We believe that a combination of more controlled genome integration methods and computational modeling will contribute to systematically optimizing the design and performance of the splicing-based memory device. Experimentally, there are many methods to integrate a plasmid carrying the whole device into specific locations in the mammalian genome with controlled copy numbers, such as recombinase-based site-directed integration of transgenes or transposon-mediated transgenesis (Oumard et al. 2006; Turan and Bode 2011; Wilson et al. 2007). Mathematical modeling of the synthetic circuit will provide information about the steps in the system that are critical for establishing a memory response.

We designed the device to be carried by one single plasmid, containing both the trigger and the reporter. The performance of the device is tunable through three variables:

the copy number of the trigger within the device plasmid, the copy number of the reporter within the plasmid, and the copy number of the whole device-carrying plasmid integrated in the mammalian genome (Figure 5-5). By varying theses three variables, we can systematically generate different designs of the device experimentally.

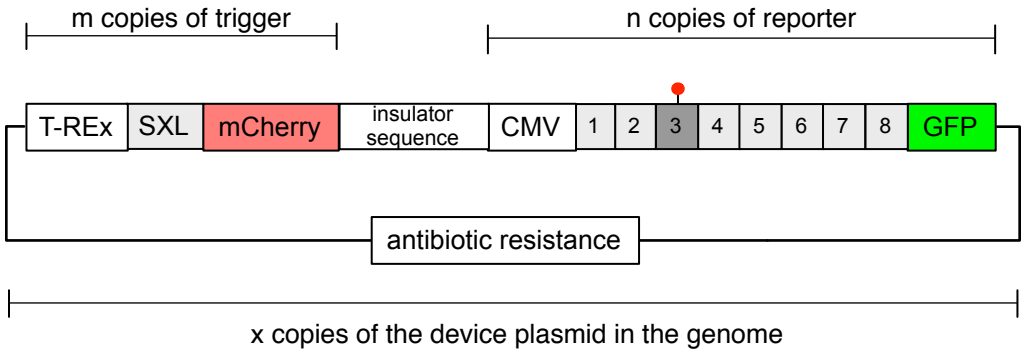

Figure 5-5: Systematic designs for controlled integration of the splicing-based device into mammalian cells.

The trigger and reporter will be carried by a single plasmid where the copy number of each part is tunable (m copies of trigger and n copies of the reporter; each copy will be separated by an insulator sequence). Depending on the experimental integration methods used, the copy number of the whole plasmid integrated can be tuned as well.

To rationally evaluate and predict the performance of the device with different designs, we applied stochastic modeling using the Gillespie algorithm (Gillespie 1977) to examine the dynamics of the device. Specifically, a preliminary model was built with simplifications described as follows:

I. The trigger activation upon dox exposure is described here in two steps: (a) A linear increase of *mCherry*-fused *Sxl* mRNA with exposure time, complemented by the degradation of the mRNA. (b) The translation of trigger Sxl mRNA to Sxl proteins. The trigger activation, based on the experimental testing, can be further modified to

detailed dox-based transcription activation, modeling the binding between dox and the Tet repressor, and the Tet repressor and the promoter. Currently, we allow 10 hours of dox exposure to activate the trigger.

II. It is estimated that a stably integrated CMV promoter generates on average 114 copies of a transcript in a single HEK-293 cell at an investigated time point (Yunger et al. 2010). Considering the strength of the CMV promoter, we define the reporter as generating a static pool of 114 *Sxl* transcripts at any moment. If needed, this simplification could be modified to model the detailed transcription activation of the CMV promoter over time when cell division is taken into account.

III. Although it is still not fully understood how Sxl regulates the splicing of exon 3 of its own transcript, it has been observed that Sxl binds to several poly-U sites upstream and downstream of exon 3 (Wang and Bell 1994; Kanaar et al. 1995). It has been suggested that splicing of exon 3 is the synergistic result of Sxl proteins that are bound to the auxiliary sites near exon 3 (Sakamoto et al. 1992; Wang and Bell 1994; Penalva et al. 1996). Additionally, it is believed that Sxl proteins cooperatively bind to the auxiliary sites in the *Sxl* pre-mRNA forming homodimers (Sakamoto et al. 1992; Wang and Bell 1994; Samuels et al. 1998). It has been suggested that these Sxl dimers are critical for carrying out the splicing decision (Louis et al. 2003). Combining the information above, we modeled the regulation of splicing of exon 3 by Sxl proteins via assuming Sxl proteins bind cooperatively to a pre-mRNA and forms dimers, as described in Louis et al. 2003. For simplicity, we only consider Sxl dimers from one binding site in the modeling but ignore the

synergistic effect from multiple Sxl dimers from many binding sites (Louis et al. 2003).

We then built a preliminary model of the splicing-based synthetic memory circuit based on the descriptions above, supposing that there is one copy of the trigger and one copy of the reporter integrated in the genome. Model variables are explained in Table 5-1. Major reactions of the model are shown as equations (1)-(13). Equations (1)-(2) depict trigger activation upon dox exposure; (3)-(5) depict the splicing activation of the reporter upon trigger activation; (6)-(9) depict the splicing-based positive feedback loop of the reporter. The parameter values used in this model are listed in Supplementary Table S5-1 (Louis et al. 2003). The initial value of each variable is set to zero except for $\tilde{S}_{reporter-mRNA}$, which is set to 114.

Table 5-1. Model variables

Model variables	Definition
S	Sxl
Dox	Exposure to dox
$S_{trigger-mRNA}$	The *mCherry* fused *Sxl* transcript produced from the trigger
$S_{trigger-protein}$	The mCherry-tagged Sxl protein produced from the trigger
$\tilde{S}_{reporter-mRNA}$	The default non-functional exon 3-included *Sxl* transcript produced from the reporter
$\tilde{S}_{reporter-mRNA} \bullet S_{trigger-protein}$	The intermediate product in splicing: an mCherry-tagged Sxl protein bound to an exon 3-included reporter *Sxl* transcript
$\tilde{S}_{reporter-mRNA} \bullet S^2_{trigger-protein}$	The intermediate product in splicing: two mCherry-tagged Sxl proteins bound to an exon 3-included reporter *Sxl* transcript as homodimers
$S_{reporter-mRNA}$	The *GFP*-fused exon 3-excluded *Sxl* transcript produced from the reporter
$S_{reporter-protein}$	The GFP-tagged Sxl protein produced from the reporter
$\tilde{S}_{reporter-mRNA} \bullet S_{reporter-protein}$	The intermediate product in splicing: a GFP-tagged Sxl protein bound to an exon 3-included reporter *Sxl* transcript
$\tilde{S}_{reporter-mRNA} \bullet S^2_{reporter-protein}$	The intermediate product in splicing: two GFP-tagged Sxl proteins bound to an exon 3-included reporter *Sxl* transcript as homodimers

(1) $Dox \xrightarrow{k1} S_{trigger-mRNA}$

(2) $S_{trigger-mRNA} \xrightarrow{k5} S_{trigger-protein}$

(3) $\tilde{S}_{reporter-mRNA} + S_{trigger-protein} \underset{l2}{\overset{k2}{\rightleftharpoons}} \tilde{S}_{reporter-mRNA} \bullet S_{trigger-protein}$

(4) $\tilde{S}_{reporter-mRNA} \bullet S_{trigger-protein} + S_{trigger-protein} \underset{l3}{\overset{k3}{\rightleftharpoons}} \tilde{S}_{reporter-mRNA} \bullet S^2_{trigger-protein}$

(5) $\tilde{S}_{reporter-mRNA} \bullet S^2_{trigger-protein} \xrightarrow{k4} S_{reporter-mRNA} + 2S_{trigger-protein}$

(6) $S_{reporter-mRNA} \xrightarrow{k5} S_{reporter-protein}$

(7) $\tilde{S}_{reporter-mRNA} + S_{reporter-protein} \underset{l2}{\overset{k2}{\rightleftharpoons}} \tilde{S}_{reporter-mRNA} \bullet S_{reporter-protein}$

(8) $\tilde{S}_{reporter-mRNA} \bullet S_{reporter-protein} + S_{reporter-protein} \underset{l3}{\overset{k3}{\rightleftharpoons}} \tilde{S}_{reporter-mRNA} \bullet S^2_{reporter-protein}$

(9) $\tilde{S}_{reporter-mRNA} \bullet S^2_{reporter-protein} \xrightarrow{k4} S_{reporter-mRNA} + 2S_{reporter-protein}$

(10) $S_{trigger-protein} \xrightarrow{d1} \varnothing$

(11) $S_{reporter-protein} \xrightarrow{d1} \varnothing$

(12) $S_{trigger-mRNA} \xrightarrow{d2} \varnothing$

(13) $S_{reporter-mRNA} \xrightarrow{d2} \varnothing$

With these initial settings of the model, the memory of transient dox exposure can be established successfully (Figure 5-6A). We then perturbed the parameters of the model as well as the initial values of $\tilde{S}_{reporter-mRNA}$ in order to investigate their effects on model performance. The results from these perturbations can determine whether a specific kinetic step of the circuit is essential for the performance of the device and contribute to the optimization of the device design. Interestingly, although this is a preliminary model, it helped us to reveal interesting insights. For example, we found that memory establishment in the model is very sensitive to the capacity of the reporter to generate Sxl transcripts, i.e. the pool of Sxl transcripts generated by the reporter that can be converted to functional female-specific isoforms and produce Sxl proteins (Figure 5-6B,C). In contrast, the rate of trigger Sxl production does not affect the model's performance on memory retention as much (data not shown). Suppose the number of transcripts generated

Figure 5-6: Stochastic modeling prediction of device performance when the production level of reporter Sxl transcripts is perturbed.

(A) Device performance predicted by stochastic modeling when the reporter can generate 114 copies of Sxl transcript. Memory of dox exposure is established. Left panel shows trigger and reporter Sxl protein productions and right panel shows functional *Sxl* mRNA (exon 3-excluded) productions from the trigger and reporter. X-axis depicts hours after dox induction and Y-axis shows number of molecules. Dox is removed after 10 hours.

(B) Device performance predicted by stochastic modeling when the reporter can generate 57 copies of Sxl transcript. Memory of dox exposure is still retained.

(C) Device performance predicted by stochastic modeling when the reporter can generate 28 copies of Sxl transcript. Memory of dox exposure is no longer retained.

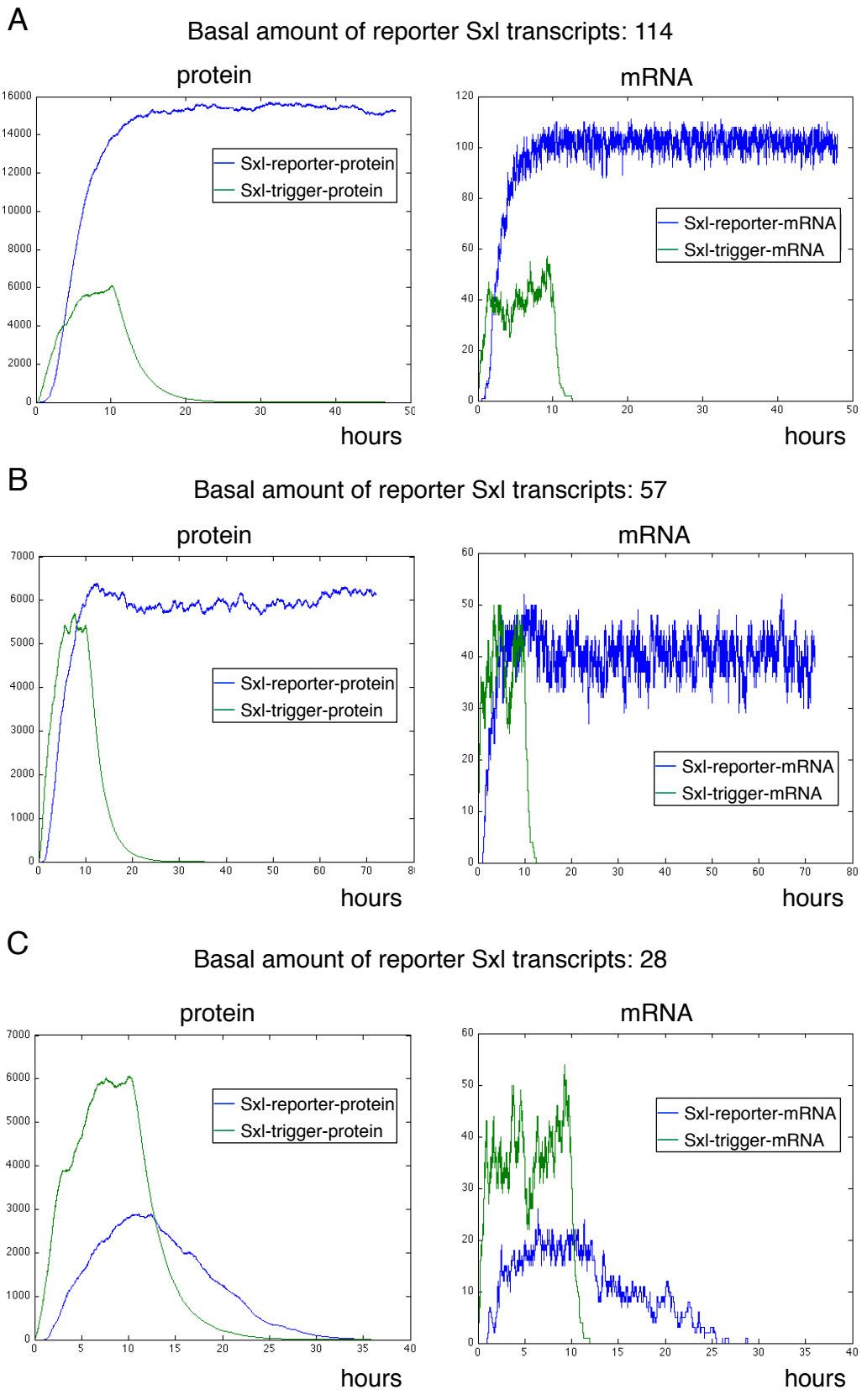

Figure 5-6 (Continued).

from a CMV promoter is fixed, the amount of reporter *Sxl* transcripts is positively correlated to the copy number of the reporter in the device. Thus, it is crucial to experimentally measure the basal transcription level of the CMV promoter in our reporter and optimize the ratio of trigger/reporter constructs in the device based on that.

DISCUSSION

Preserved AS regulation of Drosophila Sxl in mammalian cells

One surprising discovery of our study is that the self-regulation of splicing of *Drosophila Sxl* is preserved in mammalian cells, although mammalian cells do not have the gene *Sxl* nor its homolog and sex determination in mammals is profoundly different from flies. The expression of Sxl protein does not seem to have negative effects on mammalian cells (Figure 5-2A). As in *Drosophila*, Sxl transcripts containing exon 3 and surrounding intronic sequences per default generate the exon 3-inclusive male-specific isoform in mammalian cells (Figure 5-2B,D), and only the expression of a mature Sxl protein can alter the AS to generate the exon 3-exclusive, female-specific isoform (Figure 5-2B,C,D).

Although the mechanism of Sxl regulating the splicing of exon 3 is not fully understood, many factors have been identified to be involved in this process, including Spf45, *snf*, *fl(2)d*, and *vir* (Lallena et al. 2002; Salz and Flickinger 1996; Granadino et al. 1990; Hilfiker et al. 1995). All these factors are splicing regulators and unlike Sxl, have homologs in mammalian cells. Thus the machinery for Sxl to regulate its own pre-mRNA splicing is preserved in mammalian cells. This could explain why we observe that

splicing of *Drosophila Sxl* is recapitulated in mammalian cells. The observation that overexpression of *Drosophila Sxl* in U2OS cells does not lead to a phenotype could be explained by the high specificity of AS regulation by Sxl (Penalva and Sánchez 2003). So far, only *Sxl* itself, *tra* and *msl-2* in *Drosophila* have been experimentally validated as AS targets of Sxl. The splicing regulation of Sxl on these factors is involved in *Drosophila* sex determination, dosage compensation and germ line sex determination, the specific mechanisms of which are not implicated outside the genus *Drosophila* (Penalva and Sánchez 2003). As a result, the expression of foreign Sxl does not significantly affect mammalian cells.

The potential of applying Sxl auto-regulatory feedback loop in synthetic memory circuits

Our Sxl splicing-based memory device showed prompt response of the trigger to dox exposure and subsequent activation of the reporter from the trigger when stably integrated (Figure 5-4B,C). Although we have not been able to confirm whether stable establishment of memory can be achieved after removal of dox, the experimental results indicate the possibility of using the positive feedback loop of Sxl auto-splicing-regulation in memory devices.

Due to the absolute bistability in the splicing pattern of Sxl and the fast regulation of its own transcript, the auto-regulatory loop of Sxl splicing regulation has great potential for use in synthetic memory devices. Unlike most AS events, the splicing of Sxl presents a 100% switch between the two male/female AS isoforms (Penalva and Sánchez 2003). This demonstrates a tight control between the two states represented by the expression of the two splicing isoforms and low leakage in each state. In addition,

although AS is a posttranscriptional regulatory mechanism, many splicing events are coupled with transcription, in which the speed of RNA polymerase II and other activities affecting transcription can affect splicing (Pandya-Jones and Black, 2009; Listerman et al. 2006). However, it is shown that the regulation of Sxl on the splicing of exon 3 is uncoupled from transcription (Vargas et al. 2011). This means that Sxl AS is likely regulated under a fast time scale like other post-transcriptional regulations. This feature is important for building memory devices in mammalian cells as the device needs to respond to a transient trigger exposure and balance with the time scale of cell division.

The interplay of computational modeling and experimental testing in rational synthetic device design

The optimization of synthetic devices requires an efficient system to rationally test and evaluate the design of circuits. The combination of mathematical modeling and controlled genome integration methods provide such an approach.

Information about the circuit dynamics can be obtained from modeling and used to guide the design of the device (Figure 5-6; Figure 5-5). Conversely, the comparison between the prediction of the model and the experimental performance of the device will improve the modeling and increase the prediction capability of the model. We believe that close interplay between computational modeling and experimental testing will improve our understanding of the optimization of splicing-based synthetic devices, as well as the dynamics of the AS regulatory network and its potential in building synthetic devices.

MATERIALS AND METHODS

Plasmids, cell culture and stable cell line making

Female *Sxl* transcript was cloned from cDNA made from female flies. RNA from 10 females flies was extracted with Trizol (Invitrogen) and 5 μg were reverse transcribed with Superscript III (Invitrogen) and oligo-dT. Genomic regions of *Sxl* (exon 3 and adjacent intronic sequences) were PCR-amplified from *Drosophila* genomic DNA. Minigene constructs were cloned in T-Rex vector (Silver laboratory), EGFP-N1 and mCherry-N1 vector (Clontech).

U2OS cell lines (Backlow laboratory) were grown in standard conditions. Plasmid transfection was performed with LipofectAMINE 2000 (Invitrogen). To generate a staple cell line carrying the memory device through random integration, U2OS cells were co-transfected with the trigger and reporter constructs and co-selected with puromycin and hygromycin B 48 hr post-transfection. Resistant colonies were isolated with cloning cylinders (Millipore) and expanded.

RNA extraction and RT-PCR

U2OS cellular RNA was extracted by RNeasy Mini Kit (Qiagen). 1.5-3 μg RNAs were reverse-transcribed by Superscript First-Strand Synthesis System (Invitrogen). *Sxl* splicing isoforms were analyzed by RT-PCR through primers flanking the alternatively spliced region. A linear range for cDNA input was determined by running a two fold dilution series of 28 cycle PCR reactions. Products were run on 5% Tris-Borate-EDTA gels (Biorad), stained with SYBR gold stain (Invitrogen) and quantified with Image J software (NIH). Triplicate samples in the linear range were run to determine isoform

ratios. Real-time PCR primers were designed with the software Primer3 (SourceForge) and verified by electrophoresis to generate single, correctly sized products.

Western blotting

Cellular proteins were extracted with 1xRIPA buffer (IX PBS, 1% Triton X-100, 0.5% sodium deoxycholate, 0.1% SDS, protease inhibitors). 10-30 μg of proteins were loaded onto each lane of a NuPAGE 4-12% Bis-Tris gel. Proteins were transferred onto PVDF membranes and probed with GFP antibody (Aves labs) diluted in 1X PBST with 5% non-fat milk.

Flow cytometry

Cells were fixed with 4% paraformaldehype in 1X PBS and resuspended in 1X PBS. Cells were then loaded onto an LSRII (BD Biosciences) with 488 nm (DsRed) and 568 nm (FITC) lasers. Unexposed U2OS cells were used to set up levels for basal fluorophore expression. Data were analyzed with FlowJo.

ACKNOWLEDGEMENTS

We thank Sarah Saminadin-Peter for help with RNA extraction from flies; Karmella Haynes, Jeffrey Way and Daniel Ducat for thoughtful discussions; Mara Inniss for help with stable cell line making and flow cytometry analysis; Abhishek Garg for help with stochastic modeling. This work was supported by grants GM36373 and R37 from the National Institutes of Health.

REFERENCES

Ajo-Franklin CM, Drubin DA, Eskin JA, Gee EP, Landgraf D, Phillips I, Silver PA. 2007. Rational design of memory in eukaryotic cells. *Genes Dev* **21:** 2271-2276.

Atkinson MR, Savageau MA, Myers JT, Ninfa, AJ. 2003. Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in Escherichaia coli. *Cell* **113:** 597-607.

Baker BS. 1989. Sex in flies: the splice of life. *Nature* **340:** 521-524.

Bristow RG, Hill RP. 2008. Hypoxia and metabolism. Hypoxia, DNA repair and genetic instability. *Nat Rev Cancer* **8:** 180-192.

Burrill DR, Inniss MC, Boyle PM, Silver PA. 2012. Synthetic memory circuits for tracking human cell fate. *Genes Dev* **26:** 1486-1497.

Burrill DR, Silver PA. 2011. Synthetic circuit identifies sub-populations with sustained memory of DNA damage. *Genes Dev* **25:** 434-439.

Gaiano N, Fishell G. 2002. The role of notch in promoting glial and neural stem cell fates. *Annu Rev Neurosci* **25:** 471-490.

Gardner TS, Cantor CR, Collins JJ. 2000. Construction of a genetic toggle switch in *Escherichia* coli. *Nature* **403:** 339-342.

Gillespie DT. 1977. Exact stochastic simulation of coupled chemical reactions. *J phys chem* **81:** 2340-2361.

Granadino B, Campuzano S, Sánchez L. 1990. The drosophila melanogaster fl(2)d gene is needed for the female-specific splicing of Sex-lethal RNA. EMBO J **9:** 2597-2602.

Hilfiker A, Amrein H, Dübendorfer A, Schneiter R, Nöthiger R. 1995. The gene virilizer is required for female-specific splicing controlled by Sxl, the master gene for sexual development in Drosophila. *Development* **121:** 4017-4026.

Ingolia NT, Murray AW. 2007. Positive-feedback loops as a flexible biological module. *Curr Biol* **17:** 6668-677.

Kanaar R, Lee AL, Rudner DZ, Wemmer DE, Rio DC. 1995. Interaction of the sex-lethal RNA binding domains with RNA. *EMBO J* **14:** 4530-4539.

Kramer BP, Viretta AU, Daoud-EI-Baba M, Aubel D, Weber W, Fussenegger M. 2004. An engineered epigenetic transgene switch in mammalian cells. *Nat Biotechnol* **22:** 867-870.

Lallena MJ, Chalmers KJ, Llamazares S, Lamond AI, Valcárcel J. 2002. Splicing regulation at the second catalytic step by Sex-lethal involves 3' splice site recognition by SPF45. *Cell* **109:** 285-296.

Listerman I, Sapra AK, Neugebauer KM. 2006. Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells. *Nat Struct Mol Biol* **13:** 815-822.

Louis M, Holm L, Sánchez L, Kaufman M. 2003. A theoretical model for the regulation of Sex-lethal, a gene that controls sex determination and dosage compensation in Drosophila melanogaster. *Genetics* **165:** 1355-1384.

Nelson BR, Hartman BH, Georgi SA, Lan MS, Reh TA. 2007. Transient inactivation of Notch signaling synchronizes differentiation of neuronal progenitor cells. *Dev Biol* **304:** 479-498.

Oumard A, Qiao J, Jostock T, Li J, Bode J. 2006. Recommended Method for Chromosome Exploitation: RMCE-based Cassette-exchange Systems in Animal Cell Biotechnology. *Cytotechnology* **50:** 93-108.

Pandya-Jones A, Black DL. 2009. Co-transcriptional splicing of constitutive and alternative exons. *RNA* **15:** 1896-1908.

Penalva LO, Sakamoto H, Navarro-Sabaté A, Sakashita E, Granadino B, Segarra C, Sánchez L. 1996. Regulation of the gene Sex-lethal: a comparative analysis of Drosophila melanogaster and Drosophila subobscura. *Genetics* **144:** 1653-1664.

Penalva LO, Sánchez L. 2003. RNA binding protein sex-lethal (Sxl) and control of Drosophila sex determination and dosage compensation. *Microbiol Mol Biol Rev* **67:** 343-359.

Sakamoto H, Inoue K, Higuchi I, Ono Y, Shimura Y. 1992. Control of Drosophila Sex-lethal pre-mRNA splicing by its own female-specific product. *Nucleic Acids Res* **20:** 5533-5540.

Salz HK, Flickinger TW. 1996. Both loss-of-Function and gain-of-Function mutations in snf define a role for snRNP proteins in regulating Sex-lethal pre-mRNA splicing in Drosophila development. *Genetics* **144:** 95-108.

Samuels M, Deshpande G, Schedl P. 1998. Activities of the Sex-lethal protein in RNA binding and protein:protein interactions. *Nucleic Acids Res* **26:** 2625-2637.

Strogatz S. 1994. *Nonlinear dynamics and chaos: with applications to physics, biology chemistry and engineering*. Westview Press, Cambridge, MA.

Thomas R, Kaufman M. 2001. Multistationarity, the basis of cell differentiation and memory. I. Structural conditions of multistationarity and other nontrivial behavior. *Chaos* **11:** 170-179.

Turan S, Bode J. 2011. Site-specific recombinases: from tag-and-target- to tag-and-exchange-based genomic modifications. *FASEB J* **25:** 4088-4107.

Vargas DY, Shah K, Batish M, Levandoski M, Sinha S, Marras SA, Schedl P, Tyagi S. 2011. Single-molecule imaging of transcriptionally coupled and uncoupled splicing. *Cell* **147:** 1054-1065.

Vilaboa N, Fenna M, Munson, J, Roberts, SM, Voellmy, R. 2005. Novel gene switches for targeted and timed expression of proteins of interest. *Mol Ther* **12:** 290-298.

Wang J, Bell LR. 1994. The Sex-lethal amino terminus mediates cooperative interactions in RNA binding and is essential for splicing regulation. *Genes Dev* **8:** 2072-2085.

Wilson MH, Coates CJ, George AL Jr. 2007. PiggyBac transposon-mediated gene transfer in human cells. *Mol Ther* **15:** 139-145.

Xiong W, Ferrell JE Jr. 2003. A positive-feedback-based bistable 'memory module' that governs a cell fate decision. *Nature* **426:** 460-465.

Yao F, Svensjö T, Winkler T, Lu M, Eriksson C, Eriksson E. 1998. Tetracycline repressor, tetR, rather than the tetR-mammalian cell transcription factor fusion derivatives, regulates inducible gene expression in mammalian cells. *Hum Gene Ther* **9:** 1939-1950.

Yunger S, Rosenfeld L, Garini Y, Shav-Tal Y. 2010. Single-allele analysis of transcription kinetics in living mammalian cells. *Nat Methods* **7:** 631-633.

APPENDIX A: AN ALTERNATIVE SPLICING NETWORK LINKS CELL-CYCLE CONTROL TO APOPTOSIS

Reprinted from
Cell, 2010. **142**(4): 625-636.
with permission from Elsevier

Cell

An Alternative Splicing Network Links Cell-Cycle Control to Apoptosis

Michael J. Moore,[1] Qingqing Wang,[1,3] Caleb J. Kennedy,[1,3] and Pamela A. Silver[1,2,*]
[1]Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA
[2]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA, 02115, USA
[3]These authors contributed equally to this work
*Correspondence: pamela_silver@hms.harvard.edu
DOI 10.1016/j.cell.2010.07.019

SUMMARY

Alternative splicing is a vast source of biological regulation and diversity that is misregulated in cancer and other diseases. To investigate global control of alternative splicing in human cells, we analyzed splicing of mRNAs encoding Bcl2 family apoptosis factors in a genome-wide siRNA screen. The screen identified many regulators of *Bcl-x* and *Mcl1* splicing, notably an extensive network of cell-cycle factors linked to aurora kinase A. Drugs or siRNAs that induce mitotic arrest promote proapoptotic splicing of *Bcl-x*, *Mcl1*, and *caspase-9* and alter splicing of other apoptotic transcripts. This response precedes mitotic arrest, indicating coordinated upregulation of prodeath splice variants that promotes apoptosis in arrested cells. These shifts correspond to post-translational turnover of splicing regulator ASF/SF2, which directly binds and regulates these target mRNAs and globally regulates apoptosis. Broadly, our results reveal an alternative splicing network linking cell-cycle control to apoptosis.

INTRODUCTION

Nearly all human precursor messenger RNAs (pre-mRNAs) undergo alternative splicing (AS), with tremendous variation and specificity across tissues, development, and disease (Pan et al., 2008). This vast complexity is a formidable challenge to experimental and informatic analysis of AS and its physiological roles (Blencowe, 2006). One emergent concept is that functionally coherent transcript populations, termed RNA "regulons," are coregulated by dedicated RNA-binding proteins (RBPs) to promote specific biological functions (Keene, 2007). These RBPs, notably SR proteins and hnRNPs, "decode" transcript *cis* elements and control stepwise assembly of spliceosomal snRNPs: U1 at the 5′ splice -site (ss), then U2 near the 3′ ss, and finally U5/4/6 (Wahl et al., 2009; Black, 2003). Splicing control is integrated with signal transduction pathways, promoting dynamic, context-driven regulation. Collectively, these features produce robust cell- and tissue-specific signatures of exon use that shape many aspects of cell fate (Moore and Silver, 2008). Exon signatures are radically transformed in tumors, but the causes and consequences are unknown.

In this study, we examine AS of Bcl2 family apoptosis regulators *Bcl-x* and *Mcl1*. Bcl2-like proteins contain up to four Bcl2-homology (BH) domains (BH1–4) (Hardwick and Youle, 2009). Factors possessing all four BH domains, including Bcl2, Bcl-xL, and Mcl1L, antagonize apoptosis by preventing mitochondrial outer membrane permeablization (MOMP), thus sequestering proapoptotic factors in mitochondria. Factors lacking one or more BH domain, including Bid, BAD, and BAX, are proapoptotic and promote MOMP. A finely tuned balance of pro- and antiapoptotic Bcl2-like factors therefore controls mitochondrial integrity and hence downstream steps in apoptosis such as apoptosome formation and caspase activation (Wang and Youle, 2009). Remarkably, *Bcl-x*, *Mcl1*, and some other *Bcl2* family mRNAs are alternatively spliced to yield both long (L) antiapoptotic and short (S) proapoptotic forms. For *Bcl-x*, use of an alternative 5′ ss in exon 2 excludes the BH1 and BH2 domains (Akgul et al., 2004). For *Mcl1*, exon 2 skipping excludes the BH1 and BH2 domains and eliminates the downstream transmembrane domain via frame shift.

Many *cis*-regulatory elements and *trans*-acting factors exert combinatorial control of *Bcl-x* splicing. Most known regulators, including Sam68, ASF/SF2, hnRNP F/H, SRp30c, and RBM25, altered *Bcl-x* AS in vitro or when overexpressed in cell culture (Cloutier et al., 2008; Zhou et al., 2008; Paronetto et al., 2007; Garneau et al., 2005). In addition, in RNA interference (RNAi)-based loss-of-function assays, depletion of Sam68 and hnRNPA1 favored *Bcl-xL* formation, while depletion of U2 snRNP component SF3B1/SAP155 favored *Bcl-xS* (Paronetto et al., 2007; Massiello et al., 2006). Comparably little is known of *Mcl1* splicing regulation.

Beyond Bcl2-like factors, caspases, "death receptors," ligands and various adaptors are regulated by AS, suggesting broad roles in controlling apoptosis (Schwerk and Schulze-Osthoff, 2005). Many apoptosis regulators, including Bcl2-like proteins, are proto-oncogenes that contribute to apoptosis resistance in cancer (Letai, 2008; Fesik, 2005). Modulation of apoptotic factors by targeting the splicing machinery is thus an attractive strategy to facilitate tumor cell death. Furthermore, while the divergent functions of *Bcl-x* and *Mcl1* isoforms in apoptosis are well established, the physiological contexts and upstream regulation of their expression are poorly defined.

These unanswered questions illustrate a pervasive challenge in defining physiological contexts of AS regulation, because strategies for systematic evaluation of upstream regulation are limited.

Genome-scale screening of RNA regulatory events is complicated by the difficulty of visualizing RNAs in vivo, and the technical infeasibility of high-throughput measurements by RT-PCR and other methods. Splicing-sensitive fluorescent reporters are an alternative strategy that produce a robust, visual output suitable for screening efforts (Stoilov et al., 2008; Warzecha, et al., 2009; Orengo et al., 2006). Here, we present high-throughput assays that recapitulate physiological regulation of *Bcl-x* and *Mcl1* AS. In a whole-genome siRNA screen, we identified new factors that regulate the balance of anti- and proapoptotic splice isoforms, with striking enrichment for cell-cycle factors. These results define functional interactions between the cell cycle and splicing machineries in human cells that manifest in a coordinated program of AS controlling apoptosis.

RESULTS

Reporter Assays for *Bcl-x* and *Mcl1* Alternative Splicing

To develop splicing assays for high-throughput analysis, we designed splicing-sensitive reporters for the Bcl2 family apoptosis regulators *Bcl-x* and *Mcl1*. 5′ untranslated region (UTR), open reading frame (ORF), and intervening intron sequences for *Bcl-x* and *Mcl1* were cloned in C-terminal fusions with *Venus* (yellow) and *mCherry* (red) complementary DNAs (cDNAs), respectively (Figure 1A). In HeLa cells, these constructs expressed long and short spliced mRNAs at ratios similar to endogenous mRNAs (Figure 1B). To render constructs splicing-sensitive, premature termination codons (PTCs) were introduced in alternative exon regions exclusive to long splice forms. As expected, PTCs eliminated expression of long protein variants, but short forms were retained (Figure 1C).

To produce screen assay cell lines, splicing reporters were stably transfected into HeLa cells along with a constitutive mCerulean fluorescent protein (CFP) construct. All constructs used the human EF1α promoter, which contains a 5′ UTR intron, allowing dual measurements of minigene splicing and a constitutively spliced CFP reporter under identical control. To test the *Bcl-x* reporter line, we verified that siRNA-depletion of known regulator SF3B1 increased Bcl-xS-Venus expression relative to a non-targeting control (Figure 1D, left panels) (Massiello et al., 2006). Immunoblotting confirmed efficient siRNA knockdown of SF3B1, and upregulation of Bcl-xS-Venus reporter protein (Figure 1E). RT-PCR confirmed upregulation of the *Bcl-xS-Venus* mRNA (Figure 1F). Finally, RT-PCR analysis of endogenous *Bcl-x* transcript in HeLa cells verified that SF3B1 knockdown shifted splicing toward *Bcl-xS*, demonstrating congruous regulation of minigene and endogenous splicing (Figure 1G, upper panel).

We had no a priori knowledge of *Mcl1* regulators, but SF3B1 knockdown also upregulated Mcl1S-mCherry in the splicing assay, establishing a positive assay control (Figure 1D, right panels). Analysis of endogenous *Mcl1* verified this shift toward *Mcl1S* (Figure 1G, lower panel).

High-Throughput siRNA Screens for Alternative Splicing Regulators

To identify regulators of *Bcl-x* AS, >21,000 siRNA pools targeting known and predicted human genes were screened for upregulation of the *Bcl-x* reporter (Figure 2A). Three hundred and sixty-nine positive hits were identified with a Support Vector Machine (SVM) model that determined reproducibility (i.e., ''confidence'') across triplicates and signal ''strength'' relative to positive and negative control siRNAs (Figure 2B, and Figure S1 and Table S1 available online). Hits had strong gene ontology (GO) enrichments for mRNA splicing/processing, protein kinase signaling, cytoskeleton association, and cell-cycle functions (Figure 2C). Importantly, the screen blindly recovered positive control SF3B1 and several of its interactors.

For validation, hits were retested in the screen assay with four individual siRNAs from deconvoluted SMARTpools. Two hundred and seventy-four of 369 factors validated with at least one siRNA, and 160 validated with two or more (Figure 2A and Table S2). In the primary and validation screens, Bcl-x-Venus expression correlated significantly with cell death (Figures S2A and S2C). Venus showed no systematic correlation to CFP, indicating that nonspecific promoter effects were not a major source of positives (Figures S2B and S2F). In the validation screen, apoptosis was tracked by Annexin-V-Cy5 staining and showed strong coupling to proapoptotic *Bcl-x* splicing and cell death (Figures S2D and S2E). Functional enrichments, notably splicing and cell-cycle regulation, were similar between validated hits and the primary screen (Figure 2C). Subsequent analyses focused on ''high-confidence'' factors validated with two or more siRNAs. We focused specifically on aurora kinase A (AURKA) and other mitotic regulators because of their strong functional enrichment and their novelty in the context of splicing regulation.

Physiological function for screen hits was confirmed by analysis of endogenous *Bcl-x* AS in HeLa cells. siRNAs against 19 factors were tested, and 16 (~85%) significantly shifted *Bcl-x* splicing toward *Bcl-xS* (Figure S3A). The magnitude of these shifts matched or exceeded fold changes that promote apoptosis in various cell types, indicating physiological significance (Mercatante et al., 2002; Taylor et al., 1999). As further indication of this significance, AURKA knockdown strongly shifted endogenous Bcl-x protein toward Bcl-xS (Figure S3B). This shift was larger than the corresponding mRNA shift, suggesting an amplification effect during translation. Regulation of endogenous *Bcl-x* by AURKA and other hits also confirmed in MCF7 (breast adenocarcinoma) and PANC1 (pancreatic carcinoma) cells (Figure S3C). These lines differ markedly in steady state AURKA levels, with MCF7 (like HeLa) expressing high levels, and PANC1 levels closer to ''normal'' tissue (Ross et al., 2000). Thus, splicing regulation by AURKA inhibition was not restricted to HeLa cells and occurred irrespective of steady state AURKA levels.

Apoptosis Suppression Defines Direct Regulators of *Bcl-x* Alternative Splicing

In screening experiments, proapoptotic *Bcl-x* splicing correlated to apoptosis induction and increased cell death (Figure S2). However, screen hits are likely to include both direct regulators

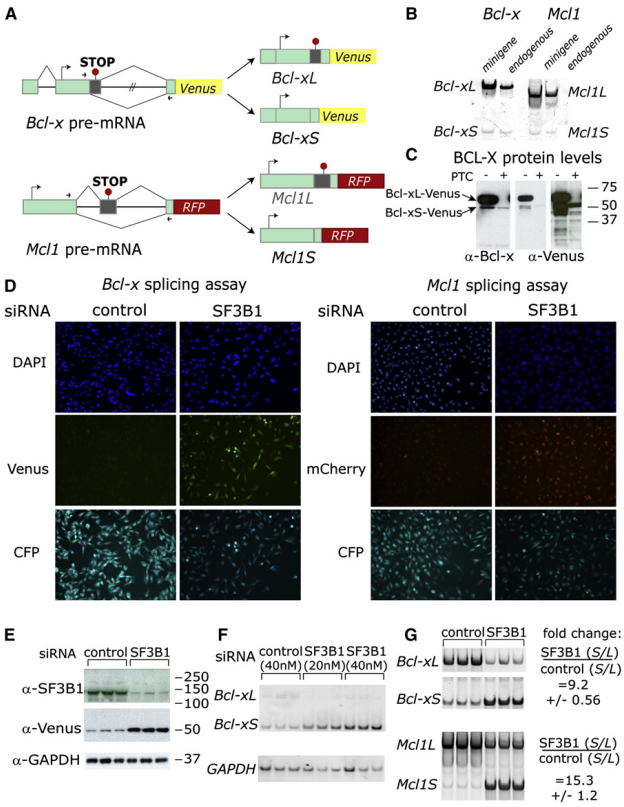

Figure 1. *Bcl-x* and *Mcl1* Alternative Splicing Reporters

(A) *Bcl-x* and *Mcl1* minigene splicing reporters are schematized. Green rectangles are exons, with dark gray as alternative regions. Black lines depict introns. Large arrows denote translation start sites, and arrowheads show primer sets used for qPCR. Inserted PTCs, exclusive to long (L) isoforms, are shown by red "stop signs."

(B) Minigene constructs are spliced like endogenous mRNAs. Splice isoforms from minigene- or mock-transfected (endogenous) HeLa cells were analyzed by RT-PCR with primer sets from (A). Because of high transient minigene expression, cDNA from minigene samples was diluted 1:10^5 relative to mock (endogenous). At this dilution, endogenous mRNAs do not contribute to products in the "minigene" lanes.

(C) A PTC introduced to the alternative exon region eliminated Bcl-xL-Venus but not Bcl-xS-Venus protein expression. HeLa cells were transfected with *Bcl-x-Venus* (PTC-) or *Bcl-x-Venus-PTC* (PTC+) constructs, and lysates were analyzed by western with indicated antibodies. Migration of isoforms is indicated to the left of images (last two lanes overexposed to show Bcl-xS-Venus).

(D) SF3B1 knockdown activated proapoptotic splicing of *Bcl-x* and *Mcl1* reporters. Stable reporter lines were transfected with SF3B1 or nontargeting siRNA pools and imaged in the indicated channels. SF3B1 knockdown increased minigene expression versus controls in both cases. SF3B1 depletion also decreased CFP modestly, possibly reflecting a role in splicing the EF1α intron. Subsequent experiments showed this effect on CFP expression was restricted to SF3B1 (see Figures S2B and S2F).

(E) Upregulation of Bcl-xS by SF3B1 depletion was confirmed by western analysis. Lysates from cells treated as in (D) were analyzed with antibodies indicated to the left of images. SF3B1 knockdown depleted SF3B1 and increased Bcl-xS-Venus relative to controls.

(F) Results in (D) and (E) were verified at the RNA level by RT-PCR. Migration of *Bcl-xS*, *Bcl-xL* (PTC causes nonsense-mediated decay), and *GAPDH* (loading control) are indicated to left of the images.

(G) Results in (D), (E), and (F) model the regulation of endogenous *Bcl-x* and *Mcl1* mRNAs. Unmodified HeLa cells were transfected with SF3B1 or control siRNAs. RT-PCR analysis confirmed that SF3B1 knockdown favored short, proapoptotic forms. *S/L* ratios are quantified to the right of the images; values are the means of three independent measurements ± SD, normalized to controls.

of *Bcl-x* splicing and factors that influence splicing indirectly via upstream activation of proapoptotic pathways (Figure 3A). To define systematically the relationship of screen hits to apoptosis, we retested all 274 validated regulators in *Bcl-x* reporter cells engineered to overexpress Bcl2. Bcl2 overexpression suppresses mitochondrial permeablization and hence canonical apoptosis progression (Kroemer, 1997). Importantly, Bcl2 directly antagonizes the function of Bcl-xS, so this strategy also suppresses potential downstream effects of Bcl-xS-Venus produced by the reporter construct. We predicted that Bcl2 overexpression would attenuate the effects of proapoptotic siRNAs that alter *Bcl-x* splicing as part of a general proapoptotic response, but not siRNAs that target more direct regulators of splicing. Supporting the latter prediction, depletion of direct regulator SF3B1 increased reporter expression despite Bcl2 overexpression (Figure 3B).

Thirty-three percent of 160 high-confidence *Bcl-x* regulators were totally suppressed by Bcl2 overexpression (Table S3). For example, Bcl2 suppressed both cell death and changes in *Bcl-x* splicing in response to depleting mRNA export factors NXF1 or NUPL2 (Figure 3C). By contrast, 67% of factors, including AURKA, showed partial or no suppression of proapoptotic *Bcl-x* splicing, despite the fact that cell counts showed efficient suppression of cell death. Thus, AURKA, not previously implicated in splicing regulation, functioned independently of Bcl2, whereas mRNA export factors, which directly interact

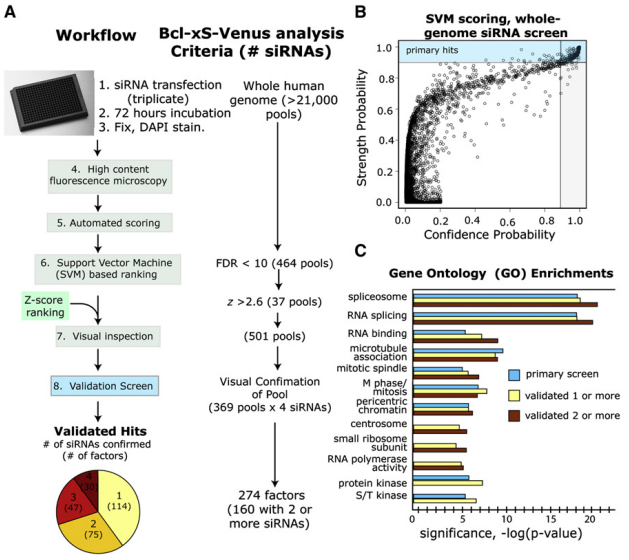

Figure 2. Whole-Genome siRNA Screen for Regulators of *Bcl-x* Alternative Splicing

(A) Screening, analysis, and hit selection are schematized. Of 369 primary hits, 274 revalidated. The pie chart indicates how many of 274 validated hits reconfirmed with one, two, three, or four siRNAs. One hundred and sixty "high-confidence" hits validated with two or more siRNAs (see the Extended Experimental Procedures for complete analysis methods).

(B) SVM probability scores are plotted for the whole-genome siRNA screen, with confidence on the x axis and strength on the y axis. The blue region shows the SVM cutoff for hit selection (see Figures S1, S2, and S3 and Tables S1 and S2 for detailed analysis and validation).

(C) GO functional enrichments from primary, revalidated (one or more siRNAs), and high-confidence (two or more siRNAs) hits are shown, plotted by relative statistical significance.

See also Figures S1, S2, and S3 and Tables S1 and S2.

with the spliceosome, were suppressed. Overall, most mRNA-splicing and cell-cycle regulators showed no or partial suppression (Figure 3D). Suppressed factors were enriched for functions in transcription, while signaling factors spanned both categories.

Discovery of *Mcl1* Alternative Splicing Regulators

To explore functional coordination of apoptotic AS events, we determined all *Bcl-x* regulators that also control functionally analogous AS of *Mcl1*. Fifty-two of 160 "high-confidence" factors regulated both events, and we validated a subset on endogenous *Mcl1* (Figure 4A, Figure S4, and Table S4). Cell-cycle functions were enriched for factors that coregulated *Bcl-x* and *Mcl1*, but not factors that regulated only *Bcl-x* (Figure 4A). Similarly, common regulators were disproportionately enriched for splicing functions versus those that only regulated *Bcl-x*. This result indicates extensive coregulation of *Bcl-x* and *Mcl1* by cell-cycle and splicing factors. In contrast, transcription regulators were enriched only for *Bcl-x*. Figure 4B schematizes hits in the *Bcl-x* and *Mcl1* screens with known or postulated functions in mRNA metabolism. Hits included "core" spliceosome components in Sm ring domains and snRNPs, as well as hnRNP and SR proteins. Interestingly, factors acting late in spliceosome assembly, including U5 snRNP components and factors previously linked to cell-cycle regulation (e.g., CDC40, CDC5L), scored positively for both *Bcl-x* and *Mcl1* (Makarova et al. 2004).

As an unbiased analysis of *Mcl1* AS, we tested siRNA pools for >700 human kinases and phosphatases in the *Mcl1* screening assay. Common regulators for *Bcl-x* and *Mcl1* included cell-cycle kinases AURKA, PLK1, and WEE1 (Figure 4C and Table S5). Further, BUB1 scored strongly for *Bcl-x* and weakly positive for *Mcl1*, though the latter did not surpass the confidence threshold. Conversely, CDK1 (CDC2) was strongly positive for *Mcl1*, and weakly so for *Bcl-x*. These data independently confirm *Mcl1* AS regulation by cell-cycle disruption.

Coupling of Cell-Cycle Control and Alternative Splicing

The screen identified many cell-cycle factors, including cancer therapy targets AURKA, PLK1, and survivin (BIRC5). Protein interaction network analysis of screen hits revealed AURKA-centered interactions spanning the cell cycle, spliceosome, and tumor suppressors (Figure 5A). As independent validation of this regulation, nocodazole, an inhibitor of microtubule polymerization, and aurora kinase inhibitors ZM447439 and VX-680 strongly induced *Bcl-xS* formation in a dose-dependent manner (Figure 5B). By contrast, the broad-spectrum kinase inhibitor staurosporine induced apoptosis but not *Bcl-xS* formation, demonstrating specific regulation by cell-cycle inhibitors. ZM447439 and VX-680 inhibit the three mammalian aurora kinases AURKA, AURKB, and germline-restricted AURKC. However, ZM447439 is ~20-fold more potent toward AURKB than AURKA in vitro, while VX-680 is ~30-fold more potent against AURKA (Ditchfield et al., 2005; Harrington et al., 2004). Since *Bcl-x* splicing regulation in screening assays was AURKA specific, VX-680 was used subsequently because of its higher specificity for AURKA. Importantly, nocodazole and VX-680 treatment also shifted endogenous *Bcl-x* and *Mcl1* splicing (Figure S5).

In flow cytometry experiments, aurora inhibitors induced mitotic arrest concomitant with upregulation of the *Bcl-x* reporter (Figure 5C). Bcl-xS-Venus levels in actively cycling G2/M cells were slightly higher than in G1, but the difference was not significant compared to arrested cells (Figure 5D). In addition, S phase

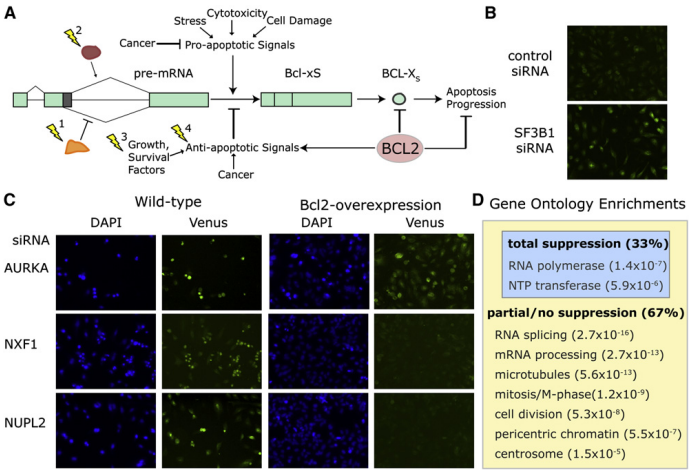

Figure 3. Apoptosis Inhibition by Bcl2 Suppressed a Subset of *Bcl-x* Regulators

(A) Screen hits (lightning bolts) may affect *Bcl-x* AS by many potential mechanisms. Regulation may be direct, such as (1) direct suppression of *Bcl-xS* formation or (2) facilitation of *Bcl-xL*. Indirect or downstream regulators may (3) promote cell survival pathways or (4) suppress proapoptotic pathways. We predicted that Bcl2 overexpression would suppress the splicing regulatory effect of siRNAs targeting indirect mechanisms, but not direct regulators.

(B) Bcl2 overexpression did not suppress the effect of SF3B1 depletion, consistent with direct regulation of *Bcl-x* splicing.

(C) Bcl2 suppressed the effect of siRNAs against mRNA export factors NXF1 and NUPL2, but not AURKA, on *Bcl-x* splicing (compare Venus images between "wild-type" and "Bcl2 overexpression"). By contrast, Bcl2 suppressed cell death in all cases (compare DAPI images).

(D) Thirty-three percent of hits were suppressed by Bcl2 overexpression, while 67% showed partial to no suppression. GO functional enrichments within these subsets are shown with p values (see Table S3).

See also Table S3.

arrest by double-thymidine block did not affect Bcl-xS-Venus levels, so the splicing response was specific to mitotic arrest (Figure 5E). Since transcription and splicing are primarily silenced in M phase (Shin and Manley, 2002), we considered two possibilities: (1) splicing changes preceded arrest or (2) a subset of cells underwent an abnormal division prior to full arrest, leading to a proapoptotic stress response. To test the latter scenario, cells were synchronized by double-thymidine block, released, and treated with VX-680. Venus fluorescence was measured 16 hr after thymidine release, thus ruling out an intervening mitotic cycle. Figure 5F shows the splicing response to VX-680 in synchronized cells was comparable to asynchronous cells. Apoptosis induction by staurosporine caused no splicing response, regardless of synchronization. These results show that stimuli that activate G2/M and spindle checkpoints trigger proapoptotic *Bcl-x* splicing, but the response precedes actual arrest.

Downstream Mechanisms of Splicing Regulation

To identify direct splicing regulators downstream of cell-cycle inhibition, we examined expression of splicing factors upon AURKA knockdown, focusing on strong screen hits or factors previously linked to *Bcl-x* splicing (Figure S6A). Among these factors, only ASF/SF2 (SFRS1)—an SR protein previously linked to cell-cycle regulation and a strong hit in the primary screen—showed specific downregulation upon AURKA knockdown (Figure 6A) (Li et al., 2005). Inhibition of AURKA with VX-680 also resulted in dose-responsive downregulation of ASF/SF2 (Figure 6B).

ASF/SF2 mRNA was not significantly reduced upon AURKA inhibition, ruling out transcriptional downregulation or effects on RNA stability (Figure S6B). However, VX-680 treatment reduced exogenous green fluorescent protein (GFP)-tagged ASF/SF2 in a manner similar to endogenous ASF/SF2, indicating posttranslational turnover as the main source of ASF/SF2 depletion (Figure 6C). AURKA inhibition also induced ASF/SF2 downregulation in apoptosis-resistant cells that overexpress Bcl2 (Figure S6C). Therefore, this regulatory pathway functions in cells where apoptosis is suppressed, consistent with our finding that AURKA knockdown shifts *Bcl-x* splicing when apoptosis is suppressed (Figure 3C).

To determine whether *Bcl-x* is a direct target of ASF/SF2, we used the CLIP method (ultraviolet [UV] crosslinking and immunoprecipitation) to analyze RNAs bound directly by ASF/SF2 in vivo. RNP complexes were covalently crosslinked in living cells by UV exposure, and ASF/SF2-RNA complexes were immunopurified (Figure 6D). RT-PCR analysis showed significant enrichment of *Bcl-x* mRNA in ASF/SF2-RNA immunoprecipitations (IPs) versus an irrelevant IgG control, along with *Mcl1* and previously known target *TPX2* (Figure 6E) (Sanford et al., 2009).

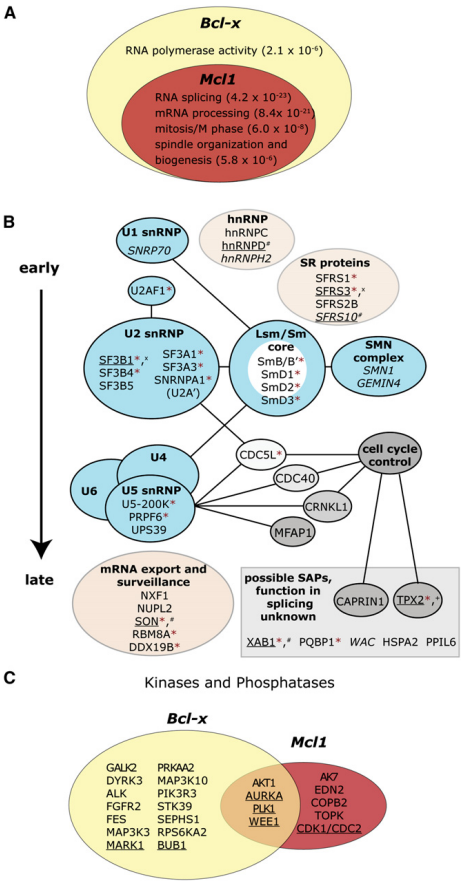

Figure 4. Coordinated Regulation of *Bcl-x* and *Mcl1* Alternative Splicing

(A) *Bcl-x* screen hits were tested for *Mcl1* regulation to identify common regulators. High-confidence hits are depicted as a Venn diagram, with GO enrichments listed with p values in the subsets where they were most significantly enriched (see Figure S4 and Table S4).

(B) Spliceosome-associated proteins (SAPs) that were *Bcl-x* screen hits are shown with established physical interactions and functional roles. Blue ellipses represent known RNP or protein complexes; pink ellipses contain functionally or structurally related factors. Low-confidence hits are italicized. Factors in gray rectangle have unknown roles in splicing but were identified in proteomic analyses or are similar to known splicing factors (Chen et al., 2007). Factors are marked as follows: *, high-confidence hit for *Mcl1*; underlined, phosphorylated in mitosis; +, putative AURKA target; x, putative CDK1 target; and #, putative PLK1 target (Dephoure et al., 2008).

(C) Hits from a kinase/phosphatase screen for *Mcl1* splicing regulation are shown, with corresponding *Bcl-x* regulators. A cutoff of $p < 10^{-5}$ in three replicates, or $p < 10^{-7}$ in two replicates was applied (see Table S5). Cell cycle factors are underlined.

See also Figure S4 and Tables S4 and S5.

APAF1 and *TNFSF13* mRNAs were detectable in input but not ASF/SF2 IPs, demonstrating that crosslinking was specific.

Our data show direct regulation of *Bcl-x* and *Mcl1* splicing by ASF/SF2 and downregulation of ASF/SF2 by AURKA inhibition. To test whether ASF/SF2 functions downstream of AURKA inhibition, we examined whether exogenous ASF/SF2 expression could "rescue" the splicing response to AURKA inhibition. *Bcl-x* reporter cells were transfected with mCherry-tagged constructs expressing wild-type (WT) ASF/SF2 or a variant lacking the second RNA recognition motif (ΔRRM2) domain. After VX-680 treatment, the *Bcl-x* splicing response in cells expressing ASF/SF2-WT was reduced ~50% relative to cells expressing ΔRRM2 or mCherry alone (Figure 6F). CLIP of exogenous GFP-tagged ASF/SF2 constructs confirmed deficient recognition of endogenous *Bcl-x* mRNA by the ΔRRM2 mutant (Figure 6G) (Cáceres et al., 1997). Therefore, enforced ASF/SF2 expression rescued the effect of AURKA inhibition on *Bcl-x* splicing, and rescue was dependent on intact splicing regulatory function.

ASF/SF2 regulates apoptosis globally in response to cell-cycle inhibition. Consistent with previous reports, siRNA depletion of ASF/SF2 activates markers of apoptosis (e.g., Annexin-V staining, DNA fragmentation, and caspase cleavage) (data not shown) (Li et al., 2005; Karni et al., 2007). To analyze apoptosis regulation by ASF/SF2 in the context of cell-cycle disruption, we depleted ASF/SF2 by RNAi and analyzed apoptosis markers upon AURKA inhibition with VX-680. ASF/SF2 depletion enhanced caspase activation and Annexin-V staining relative to a nontargeting siRNA control (Figures 7A and 7B). In addition, we compared the effects of ASF/SF2 depletion to the depletion of two splicing factors (SRPK1 and SNRP70) that had no significant effect on endogenous *Bcl-x* splicing and a third factor (Sam68) that has antagonistic effects to ASF/SF2 on *Bcl-x* splicing (Figure S3A and Table S1) (Paronetto et al., 2007). ASF/SF2 depletion sensitized cells to VX-680-induced apoptosis relative to SRPK1, SNRP70, or Sam68 depletion. In contrast, ASF/SF2 depletion did not enhance staurosporine-induced apoptosis relative to SRPK1, SNRP70, or Sam68 depletion (Figure S6D). We could not reliably analyze the effects of enforced ASF/SF2 expression on apoptotic markers because overexpression was toxic in the time frame required for this analysis. In sum, ASF/SF2 loss-of-function triggers apoptosis and specifically sensitizes cells to apoptosis induced by AURKA inhibition.

To examine broader functions of AS in this apoptotic response, we analyzed *caspase-9* (*CASP9*) and *caspase-2* (*CASP2*) mRNAs, which produce short (S), antiapoptotic, or long (L), proapoptotic isoforms (Schwerk and Schulze-Osthoff, 2005). Depletion of ASF/SF2 and AURKA favored proapoptotic *CASP9* splicing (Figure 7C) (Massiello and Chalfant, 2006). In addition, ASF/SF2 depletion favored proapoptotic *CASP2* splicing, although AURKA knockdown had no significant effect. We extended our analysis to additional factors with diverse apoptotic functions and AS patterns, including exon skipping, alternative 5′ ss use, and intron retention. Splice isoforms were measured with an isoform-sensitive qPCR strategy upon

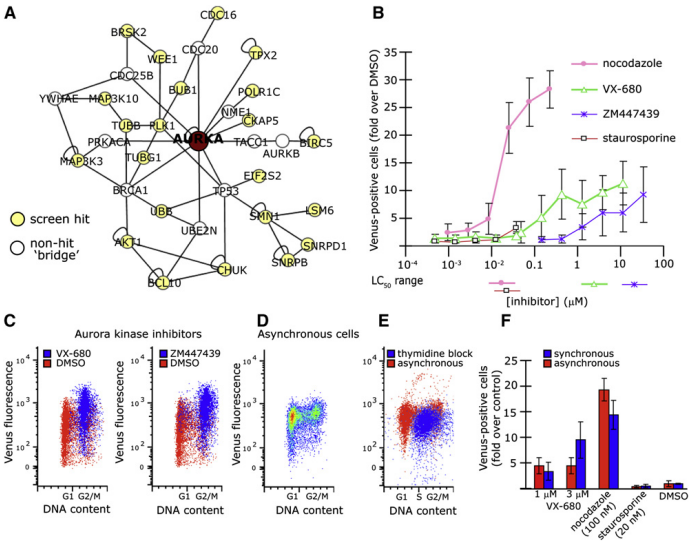

Figure 5. Alternative Splicing Regulation Is Coupled to Cell-Cycle Control

(A) Screen hits span an AURKA-centered protein interaction network. Circles ("nodes") depict proteins, connected by lines ("edges") representing validated protein-protein interactions. A maximum of two nonhit bridges was permitted to connect hits to AURKA, as described in the Extended Experimental Procedures.

(B) Drugs disrupting the cell-cycle promote proapoptotic *Bcl-x* AS. *Bcl-x* reporter cells were treated with nocodazole, aurora inhibitors VX-680 or ZM447439, or staurosporine for 18 hr. 3-fold dilution series were tested with the following maximum concentrations: 200 nM nocodazole, 10 μM VX-680, 30 μM ZM447439, and 33 nM staurosporine. Percent-Venus-positive values for drug treatments were normalized to DMSO treatments; plotted values reflect means of 8–12 measurements ± SD. Nocodazole, VX-680, and ZM447439, but not staurosporine, caused significant proapoptotic shifts in *Bcl-x* splicing ($p < 0.001$). LD_{50} ranges derived from cell counts are shown below the x axis. Wells with >33 nM staurosporine had >95% death, so they were not quantified (see Figure S5).

(C) Aurora inhibitors induced mitotic arrest coupled with proapoptotic *Bcl-x* splicing. *Bcl-x* reporter cells were treated with 3 μM VX-680, 10 μM ZM447439, or DMSO, stained with propidium iodide, and analyzed for DNA content and Venus fluorescence by flow cytometry.

(D) Asynchronous, actively cycling cells from (C) show little to no *Bcl-x* splicing fluctuations relative to arrested cells.

(E) S phase arrest by double thymidine block did not alter *Bcl-x* splicing, indicating specificity for mitotic arrest. Cells were analyzed as in (C).

(F) *Bcl-x* splicing regulation precedes mitotic arrest. Asynchronous or thymidine-synchronized cells from (E) were treated with the indicated inhibitors or DMSO. Cells were analyzed by automated microscopy 16 hr after thymidine release, precluding an intervening round of mitosis for synchronized cells. The response in presynchronized cells confirms a splicing shift upstream of mitotic arrest.

See also Figure S5.

knockdown of various screen hits, revealing widespread changes in splicing upon silencing of cell-cycle regulators (Figures 7D). This approach is only semiquantitative, so we focused on identifying biologically robust events by Z score analysis of four independent experiments (Figures S7A and S7B). Hierarchical clustering of knockdown profiles revealed overlapping regulation by cell-cycle kinases AURKA, CDC2, and BUB1. In addition, splicing factor profiles grouped by validated spliceosomal interactions; the U2 snRNP factors SF3B1 and SF3A1 overlapped significantly, as did U5 factors PRPF6 and U5-200K. Interestingly, ASF/SF2 targets clustered with AURKA and kinetochore component NDC80, suggesting overlapping regulation in this specific set of mRNAs. As confirmation of direct regulation, binding of ASF/SF2 to selected targets *CASP9*, *CASP2*, *ATG4*, and *BMF* was established by CLIP (Figure S7D).

DISCUSSION

Genome-wide Identification of Alternative Splicing Regulators in Human Cells

We developed a strategy for systematic, high-throughput analysis of AS regulation in human cells. Our siRNA screens identified many spliceosomal regulators of *Bcl-x* and *Mcl1* splicing, as well as functional connections to cell-cycle control, cytoskeleton dynamics, and signal transduction (Figure 2C). Strong functional enrichment for known splicing factors was an encouraging indication of specificity. In addition, a high validation rate of screen hits on endogenous AS indicated faithful modeling of physiological regulation in our reporter assays (Figure S3). Broadly, our efforts expand the catalog of factors linking mRNA metabolism and apoptosis and provide a platform for

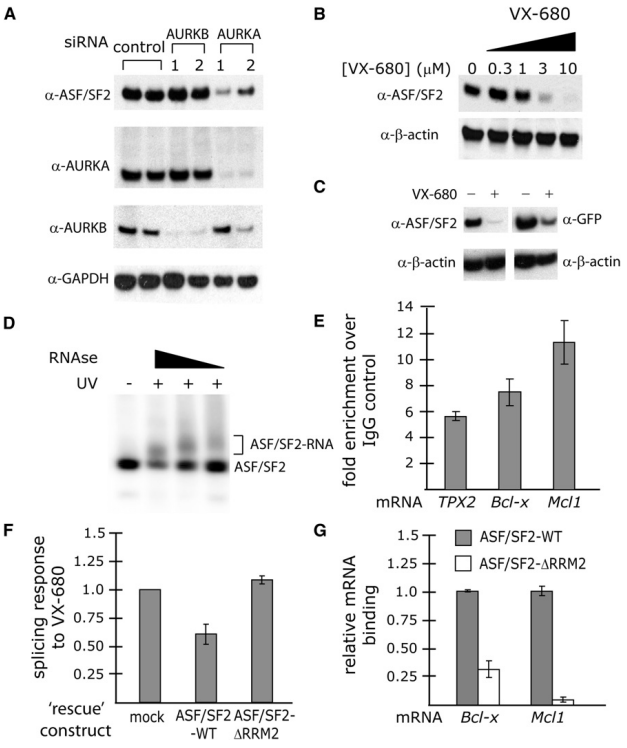

Figure 6. ASF/SF2 Regulates Splicing Downstream of AURKA

(A) siRNA depletion of AURKA downregulated ASF/SF2. Cells were transfected with two different siRNAs for AURKA or AURKB, or control siRNAs. Lysates were analyzed by western analysis for indicated factors. No other splicing factors measured in these samples showed significant changes (Figure S6A), indicating specificity for ASF/SF2.

(B) VX-680 treatment downregulated ASF/SF2 in a dose-dependent manner.

(C) ASF/SF2 is downregulated posttranslationally upon AURKA inhibition (see also Figure S6B). HeLa cells were transfected with a GFP-ASF/SF2 construct, and then treated with VX-680 or DMSO. Western analysis showed that VX-680 downregulated exogenous ASF/SF2, indicating posttranslational turnover. ASF/SF2 downregulation also occurred in apoptosis-resistant cells overexpressing Bcl2 (see Figure S6C).

(D) Visualization of ASF/SF2-RNA complexes by CLIP is shown. RNP complexes were UV cross-linked in live HeLa cells, ASF/SF2 was immunopurified, and RNA was end labeled with γ-^{32}P-ATP. Complexes were run on SDS-PAGE and visualized by autoradiography. Noncrosslinked ASF/SF2 (UV− lane) was also labeled, as observed in Sanford et al. (2009).

(E) ASF/SF2 directly binds *Bcl-x* and *Mcl1* mRNAs. RT-PCR analysis of ASF/SF2-bound RNAs showed significant enrichment of *Bcl-x*, *Mcl1*, and known target *TPX2* versus an irrelevant IgG control. Values are the means of three independent measurements ± SD.

(F) Exogenous expression of ASF/SF2 attenuated the effect of AURKA inhibition on *Bcl-x* splicing. *Bcl-x* reporter cells were transfected with mCherry-tagged ASF/SF2, a mutant lacking the second RRM domain (ΔRRM2), or mCherry alone. Eighteen hours after transfection, cells were treated with 3 μm VX-680 or DMSO for 24 hr. The fraction of Venus-positive cells among mCherry-positive cells was determined by automated microscopy. VX-680-treated samples were background corrected by subtraction of values from DMSO treatment. Values are the means of three separate measurements ± SD, expressed relative to the mCherry-transfected control. ASF/SF2-WT attenuated the *Bcl-x* response to VX-680 by ~50% versus ΔRRM2. CFP expression was not affected, indicating specific attenuation of *Bcl-x* splicing.

(G) ASF/SF2-ΔRRM2 shows deficient association with *Bcl-x* and *Mcl1* mRNAs. CLIP analysis of exogenous, GFP-tagged ASF/SF2-WT or -ΔRRM2 was performed with α-GFP antisera. Values are the means of three independent measurements ± SD.

See also Figure S6.

future high-throughput analyses of mRNA processing in living cells.

Screen hits spanned essential processes, such as transcription, cell division, and prosurvival signaling, with extensive mechanistic links to apoptosis, implicating splicing regulation in diverse proapoptotic pathways. The strong correlation of *Bcl-xS* formation with apoptosis induction and cell death further support a physiological function for splicing regulation in apoptosis (Figure S2). To investigate the functional connection between splicing regulation and apoptosis, we retested factors in cells rendered resistant to apoptosis by exogenous Bcl2 expression (Figure 3). Surprisingly, cell-cycle factors, like most spliceosomal factors, were generally not affected by Bcl2 expression. Therefore, functional coupling of cell-cycle disruption to proapoptotic splicing remained intact in cells in which the canonical apoptosis pathway was suppressed. This finding suggests that coordinated, proapoptotic splicing could be exploited to target apoptosis-resistant cell populations, such as tumors.

Our screening experiments also revealed functional coordination of proapoptotic splicing events. Fifty-two of 160 "high-confidence" *Bcl-x* regulators affected functionally analogous splicing of *Mcl1*, and common regulators were disproportionately enriched for splicing and cell-cycle functions (Figure 4A). Overlapping spliceosomal regulators of *Bcl-x* and *Mcl1* included "alternative splicing factors" (e.g., hnRNPs and SR proteins) and "core" spliceosome components spanning most steps in spliceosome assembly (Figure 4B). Interestingly, U5 snRNP

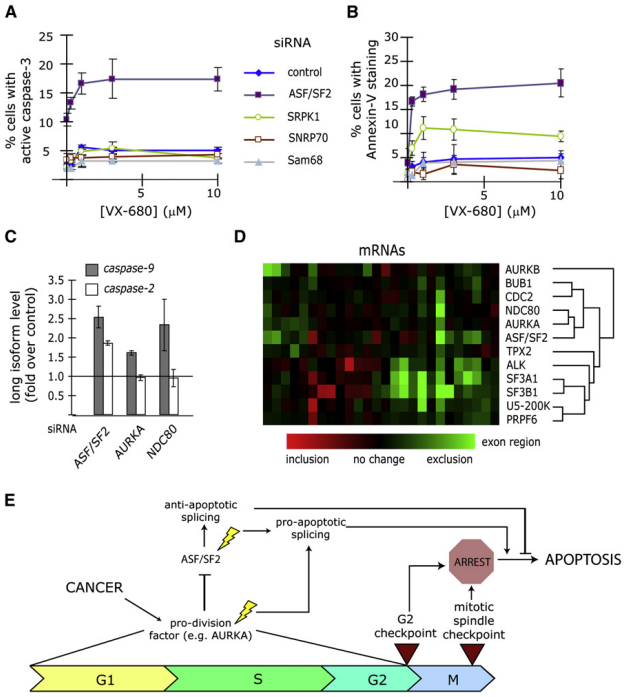

Figure 7. An Apoptotic Alternative Splicing Program Linked to Cell-Cycle Control

(A) ASF/SF2 downregulation triggers apoptosis and sensitizes cells to apoptosis induction by AURKA inhibition. HeLa cells were transfected with indicated siRNAs, incubated to allow factor depletion, and then treated with varying concentrations of VX-680. Apoptosis was measured via cleavage of a fluorescent caspase-3 substrate. Depletion of ASF/SF2, but not SRPK1, SNRP70, or Sam68, sensitized cells to VX-680-induced apoptosis. This selective sensitization by ASF/SF2 depletion was not observed for staurosporine-induced apoptosis (see Figure S6D). Data reflect the means of three independent transfections ± SD.

(B) Analysis in (A) was repeated with Annexin-V staining used as the apoptosis marker.

(C) *CASP9* and *CASP2* AS were analyzed in HeLa cells by qPCR after transfection with the indicated siRNAs. Relative levels of proapoptotic isoform (L) normalized to negative controls are shown. Values are the means of three independent measurements ± SD.

(D) Results of AS profiling by real-time qPCR are shown in a heat-map of Z scores averaged across four biological replicates, along with unsupervised hierarchical clustering (see Figure S7 for an expanded view and validation of these data).

(E) A model of cell-cycle arrest coupled to apoptosis via AS is shown. We propose anticipatory, proapoptotic splicing decisions in interphase that later promote apoptosis in arrested cells. Lightning bolts represent siRNAs, drugs, or other stimuli that disrupt the cell cycle and promote proapoptotic splicing of *Bcl-x*, *Mcl1*, *CASP9*, and other targets.

See also Figures S6 and S7.

and associated factors that function late in spliceosome assembly affected splice-site choice for *Bcl-x*, *Mcl1*, and other mRNAs (Figures 4B, 7C, and 7D). Current mechanistic models support splice-site pairing early in spliceosome assembly upon U2 snRNP recruitment (Dassah et al., 2009). Our findings suggest that perturbations of late steps in spliceosome function may promote remodeling of splice-site choice downstream of the rate-limiting steps that normally determine splicing decisions.

Cell-Cycle Disruption Coupled to Proapoptotic Alternative Splicing

Our screen for AS regulators identified factors involved in mitotic spindle assembly, kinetochore and centromere dynamics, microtubule transport, centrosome duplication and dynamics, and cell division (Figure 5A). Well-characterized pharmacological inhibitors, notably for AURKA, confirmed functional coupling of cell-cycle and splicing regulation (Figure 5B and Figure S5). In flow cytometry experiments, proapoptotic *Bcl-x* splicing was coupled to mitotic arrest, but not S phase arrest or G2/M in actively cycling cells (Figures 5C–5E). Since prevailing evidence supports suppression of transcription and splicing in mitosis, it is unlikely that the observed response initiates in mitotically arrested cells (Shin and Manley, 2002). Moreover, when cells were presynchronized to preclude an intervening cell division, VX-680 still induced a robust splicing response (Figure 5F). Therefore, this splicing regulation precedes mitotic arrest, indicating accumulation of prodeath factors in anticipation of G2/M and/or mitotic spindle checkpoints. Since cell-cycle disruption also promoted a proapoptotic switch in *CASP9* splicing and altered splicing of other apoptotic regulators, this regulation likely extends to many coregulated mRNAs (Figures 7C and 7D).

Many previous studies have linked cell-cycle and splicing regulation, reaching back to the seminal isolation of cell-division cycle (CDC) loci in *S. cerevisiae* that encode spliceosome components, including factors (CDC5L and CDC40) identified here. Recent genetic and proteomic analyses of mitotic spindle assembly have reasserted the notion that the cell-cycle and splicing machineries are intertwined, but functional consequences have not been established (Björklund et al., 2006, Rines et al., 2008). We observe coordinated proapoptotic splicing upon inhibition of key cell-cycle factors, but preceding mitotic arrest, indicating anticipatory upregulation of prodeath factors that may subsequently propagate apoptosis in arrested cells. Since arrested cells cannot mount de novo transcriptional and

splicing responses, this finding suggests a genetic mechanism for cell-cycle checkpoint response.

ASF/SF2 Regulates an Apoptotic Splicing Network

We present evidence that ASF/SF2 downregulation mediates a proapoptotic splicing response to cell-cycle disruption. ASF/SF2 was a hit in the *Bcl-x* and *Mcl1* screens, and its overexpression was previously shown to favor *Bcl-xL* formation (Paronetto et al., 2007). AURKA inhibition triggered posttranslational turnover of ASF/SF2, but not other SR proteins or splicing factors identified in the screen (Figures 6A–6C and Figure S6A). In addition, ASF/SF2 directly binds *Bcl-x* and *Mcl1* mRNAs, and its enforced expression attenuated the effect of AURKA inhibition on *Bcl-x* splicing (Figures 6D–6G).

ASF/SF2 regulation extends beyond *Bcl-x* to global regulation of apoptosis (Li et al., 2005). ASF/SF2 depletion, but not depletion of several splicing factors that were not validated screen hits, sensitized cells to apoptosis induction by AURKA inhibition (Figures 7A and 7B). In contrast, ASF/SF2 depletion did not specifically enhance apoptosis in response to staurosporine (Figure S6D). ASF/SF2-mediated splicing regulation extended to other functional targets, including *CASP9*, which underwent a proapoptotic splicing shift upon depletion of AURKA and ASF/SF2 (Figure 7C). Our analysis of some 25 additional AS targets revealed overlapping effects of AURKA and ASF/SF2 inhibition (Figure 7D). The fact that some mRNAs (e.g., *CASP2*) did not respond identically to AURKA and ASF/SF2 knockdown may reflect differential effects of targeting upstream versus direct regulators of splicing, or the activity of additional, unidentified splicing regulators. Nonetheless, these experiments strongly indicate that ASF/SF2 downregulation globally shifts alternative splicing patterns to promote apoptosis upon disruption of the cell cycle.

Alternative Splicing in a Global Context

Our data link coordinated proapoptotic splicing to activated cell-cycle checkpoints (Figure 7E). These findings are relevant to basic mechanisms of apoptosis in arrested cells, which are poorly defined, and the exploitation of these pathways in cancer therapy. Aberrant expression of AURKA and other prodivision factors promotes spindle checkpoint abrogation, chromosomal instability, and aneuploidy in tumors (Lapenna and Giordano, 2009). Our findings indicate that AURKA activation may also propagate oncogenesis by promoting antiapoptotic splicing of crucial cell-death regulators. Conversely, AURKA inhibition elicits parallel upregulation of prodeath variants, which may contribute to its demonstrated efficacy as a cancer drug target. Importantly, this regulation remains intact in apoptosis-resistant cells, suggesting functional and potential therapeutic relevance in tumors.

ASF/SF2 downregulation triggers proapoptotic splicing and promotes apoptosis. Importantly, ASF/SF2 is a proto-oncogene in its own right, and its depletion is linked to genome instability and G2/M arrest (Karni et al., 2007; Li and Manley, 2005). Our data highlight RNA targets that may contribute to its transforming properties. Taken together, these findings suggest that ASF/SF2 has multiple, synergistic functions in controlling cell arrest and apoptosis, underscoring its role as a central regulator of mRNA metabolism in cancer. Finally, it bears reiterating that our systematic screening approach was the original source of these biological insights, as well as many new regulators and functional connections not expounded here. We believe future screening efforts will be instrumental in defining global roles for AS in physiological and disease contexts.

EXPERIMENTAL PROCEDURES

siRNA Screening

Cells were transfected in triplicate in 384-well plates with >21,000 siRNA SMARTpools (Thermo-Fisher) for known and predicted human genes. Seventy-two hours later, cells were fixed, DAPI stained, and imaged on the Image Xpress Micro microscope (Molecular Devices). Cell counts (DAPI), percent of Venus-positive cells, and average per-cell CFP signal were determined with MetaXpress software. Wells were then assigned probabilistic "strength" and "confidence" scores in the Venus channel using a support vector machine (SVM) model trained on positive and negative control data (Figure S1, Table S1, and the Extended Experimental Procedures). For independent verification, plate-wide, median-normalized Z scores were also tabulated. Wells with SVM strength scores with false discovery rate (FDR) <10, *or* z > 2.6 were inspected visually.

Three hundred and sixty-nine positives were retested in a validation screen with four siRNAs from deconvoluted SMARTpools. siRNAs >2 standard deviations (SDs) above negative controls in the Venus channel were deemed positive (Table S2). Bcl2-overexpression and *Mcl1* screens were performed and analyzed as for the validation screen (Tables S3 and S4).

RNA Extraction and RT-qPCR

Cellular RNA was extracted with Trizol (Invitrogen) reagent, and 2.5–5 μg were reverse transcribed with Superscript III (Invitrogen) and oligo-dT. *Bcl-x*, *Mcl1*, *CASP9*, and *CASP2* isoforms were analyzed by qPCR with primers flanking alternative exon sequences. For each product, a linear range of cDNA input was determined empirically by running of a 2-fold dilution series in 26–28 cycle reactions. Products were run on 5% TBE-PAGE gels, stained with Sybr-Gold (Invitrogen), and quantified with QuantityOne software (Biorad).

Real-time qPCR primer sets were designed with Primer3 webware and tested by electrophoresis to verify correctly sized products (Table S6). Reactions were prepared with Sybr-Green PCR mix (ABI), run on the ABI 7900HT system, and quantified by a ΔΔC(t) method. Amplicons from alternative regions were first normalized to a constitutive amplicon from the same transcript. To correct for differential primer efficiencies, data were then normalized to negative controls. Z scores were tabulated across four independent experiments and clustered with Eisen Cluster 2.0.

Splicing Assays with Drug Treatments

Sources for inhibitors are in the Extended Experimental Procedures. Cells were treated in 96-well plates for 18–24 hr and analyzed by automated microscopy. For flow cytometry, cells were fixed in 3% PFA/1X PBS then 70% ethanol, permeablized in 1X PBS/0.1% Triton X-100, and stained with 20 μg/ml propidium iodide. DNA content and Venus fluorescence were analyzed on the LSRII system (BD Biosciences).

Western Blotting

Proteins were extracted in RIPA buffer (IX PBS, 1% Triton X-100, 0.5% sodium deoxycholate, 0.1% SDS, protease inhibitors) and quantified by Protein Dye-Binding Assay (Biorad). Ten to fifty micrograms of total protein were run on 4%–12% Bis-Tris polyacrylamide gels (Invitrogen), transferred to PVDF membranes, and probed with antibodies listed in Supplemental Data.

CLIP

UV crosslinking and RNP purification/visualization were performed as described (Sanford et al., 2009). For quantification, CLIP and input samples were treated with Proteinase K (Ambion), and RNA was extracted, reverse transcribed, and analyzed by real-time qPCR.

Apoptosis Assays

Annexin-V-Cy5 (Abcam) staining was quantified by automated microscopy. Caspase activation was measured with the DEVD-Nucview 488 reagent (Biotum), which stains nuclei specifically after it is cleaved by caspase-3. Cells were incubated for 1 hr with 2.5 μM substrate, fixed, and analyzed by automated microscopy.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and six tables and can be found with this article online at doi:10.1016/j.cell.2010.07.019.

ACKNOWLEDGMENTS

We thank Junying Yuan and Robin Reed for reagents, Jodene Moore for help with flow cytometry, and Jessica Hurt, Natalie Gilks-Farny, Daniel Ducat, William Senapedis, Faisal Aldaye, Deborah Flusberg, and Ian Swinburne for thoughtful critiques and discussions. RNAi screening capability was provided by the ICCB-Longwood screening facility at Harvard Medical School. We are indebted to ICCB-L staff Caroline Shamu, Sean Johnston, Stewart Rudnicki, Zac Cooper, and David Wrobel. M.J.M. was supported by funding from the National Science Foundation. Work was supported by grant NIHGM057476 to P.A.S. from the National Institutes of Health.

Received: September 3, 2009
Revised: February 18, 2010
Accepted: June 15, 2010
Published online: August 12, 2010

REFERENCES

Akgul, C., Moulding, D.A., and Edwards, S.W. (2004). Alternative splicing of Bcl-2-related genes: functional consequences and potential therapeutic applications. Cell. Mol. Life Sci. *17*, 2189–2199.

Björklund, M., Taipale, M., Varjosalo, M., Saharinen, J., Lahdenperä, J., and Taipale, J. (2006). Identification of pathways regulating cell size and cell-cycle progression by RNAi. Nature *439*, 1009–1013.

Black, D.L. (2003). Mechanisms of alternative pre-messenger RNA splicing. Annu. Rev. Biochem. *72*, 291–336.

Blencowe, B.J. (2006). Alternative splicing: new insights from global analyses. Cell *14*, 37–47.

Cáceres, J.F., Misteli, T., Screaton, G.R., Spector, D.L., and Krainer, A.R. (1997). Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity. J. Cell Biol. *138*, 225–238.

Chen, Y.I., Moore, R.E., Ge, H.Y., Young, M.K., Lee, T.D., and Stevens, S.W. (2007). Proteomic analysis of in vivo-assembled pre-mRNA splicing complexes expands the catalog of participating factors. Nucleic Acids Res. *35*, 3928–3944.

Cloutier, P., Toutant, J., Shkreta, L., Goekjian, S., Revil, T., and Chabot, B. (2008). Antagonistic effects of the SRp30c protein and cryptic 5′ splice sites on the alternative splicing of the apoptotic regulator Bcl-x. J. Biol. Chem. *283*, 21315–21324.

Dassah, M., Patzek, S., Hunt, V.M., Medina, P.E., and Zahler, A.M. (2009). A genetic screen for suppressors of a mutated 5′ splice site identifies factors associated with later steps of spliceosome assembly. Genetics *182*, 725–734.

Dephoure, N., Zhou, C., Villén, J., Beausoleil, S.A., Bakalarski, C.E., Elledge, S.J., and Gygi, S.P. (2008). A quantitative atlas of mitotic phosphorylation. Proc. Natl. Acad. Sci. USA *105*, 10762–10767.

Ditchfield, C., Keen, N., and Taylor, S.S. (2005). The Ipl1/Aurora kinase family: methods of inhibition and functional analysis in mammalian cells. Methods Mol. Biol. *296*, 371–381.

Fesik, S.W. (2005). Promoting apoptosis as a strategy for cancer drug discovery. Nat. Rev. Cancer *5*, 876–885.

Garneau, D., Revil, T., Fisette, J.F., and Chabot, B. (2005). Heterogeneous nuclear ribonucleoprotein F/H proteins modulate the alternative splicing of the apoptotic mediator Bcl-x. J. Biol. Chem. *280*, 22641–22650.

Hardwick, J.M., and Youle, R.J. (2009). SnapShot: BCL-2 proteins. Cell *138*, 404, 404.e1.

Harrington, E.A., Bebbington, D., Moore, J., Rasmussen, R.K., Ajose-Adeogun, A.O., Nakayama, T., Graham, J.A., Demur, C., Hercend, T., Diu-Hercend, A., et al. (2004). VX-680, a potent and selective small-molecule inhibitor of the Aurora kinases, suppresses tumor growth in vivo. Nat. Med. *10*, 262–267.

Karni, R., de Stanchina, E., Lowe, S.W., Sinha, R., Mu, D., and Krainer, A.R. (2007). The gene encoding the splicing factor SF2/ASF is a proto-oncogene. Nat. Struct. Mol. Biol. *14*, 185–193.

Keene, J.D. (2007). RNA regulons: coordination of post-transcriptional events. Nat. Rev. Genet. *8*, 533–543.

Kromer, G. (1997). The proto-oncogene Bcl-2 and its role in regulation apoptosis. Nat. Med. *3*, 614–620.

Lapenna, S., and Giordano, A. (2009). Cell cycle kinases as therapeutic targets for cancer. Nat. Rev. Drug Discov. *8*, 547–566.

Letai, A.G. (2008). Diagnosing and exploiting cancer's addiction to blocks in apoptosis. Nat. Rev. Cancer *8*, 121–132.

Li, X., and Manley, J.L. (2005). Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. Cell *122*, 365–378.

Li, X., Wang, J., and Manley, J.L. (2005). Loss of splicing factor ASF/SF2 induces G2 cell cycle arrest and apoptosis, but inhibits internucleosomal DNA fragmentation. Genes Dev. *19*, 2705–2714.

Makarova, O.V., Makarov, E.M., Urlaub, H., Will, C.L., Gentzel, M., Wilm, M., and Lührmann, R. (2004). A subset of human 35S U5 proteins, including Prp19, function prior to catalytic step 1 of splicing. EMBO J. *23*, 2381–2391.

Massiello, A., and Chalfant, C.E. (2006). SRp30a (ASF/SF2) regulates the alternative splicing of caspase-9 pre-mRNA and is required for ceramide-responsiveness. J. Lipid Res. *47*, 892–897.

Massiello, A., Roesser, J.R., and Chalfant, C.E. (2006). SAP155 Binds to ceramide-responsive RNA cis-element 1 and regulates the alternative 5′ splice site selection of Bcl-x pre-mRNA. FASEB J. *20*, 1680–1682.

Mercatante, D.R., Mohler, J.L., and Kole, R. (2002). Cellular response to an antisense-mediated shift of Bcl-x pre-mRNA splicing and antineoplastic agents. J. Biol. Chem. *277*, 49374–49382.

Moore, M.J., and Silver, P.A. (2008). Global analysis of mRNA splicing. RNA *14*, 197–203.

Orengo, J.P., Bundman, D., and Cooper, T.A. (2006). A bichromatic fluorescent reporter for cell-based screens of alternative splicing. Nucleic Acids Res. *34*, e148.

Pan, Q., Shai, O., Lee, L.J., Frey, B.J., and Blencowe, B.J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat. Genet. *40*, 1413–1415.

Paronetto, M.P., Achsel, T., Massiello, A., Chalfant, C.E., and Sette, C. (2007). The RNA-binding protein Sam68 modulates the alternative splicing of Bcl-x. J. Cell Biol. *176*, 929–939.

Rines, D.R., Gomez-Ferreria, M.A., Zhou, Y., DeJesus, P., Grob, S., Batalov, S., Labow, M., Huesken, D., Mickanin, C., Hall, J., et al. (2008). Whole genome functional analysis identifies novel components required for mitotic spindle integrity in human cells. Genome Biol. *9*, R44.

Ross, D.T., Scherf, U., Eisen, M.B., Perou, C.M., Rees, C., Spellman, P., Iyer, V., Jeffrey, S.S., Van de Rijn, M., Waltham, M., et al. (2000). Systematic variation in gene expression patterns in human cancer cell lines. Nat. Genet. *3*, 227–235.

Sanford, J.R., Wang, X., Mort, M., Vanduyn, N., Cooper, D.N., Mooney, S.D., Edenberg, H.J., and Liu, Y. (2009). Splicing factor SFRS1 recognizes a functionally diverse landscape of RNA transcripts. Genome Res. *19*, 381–394.

Schwerk, C., and Schulze-Osthoff, K. (2005). Regulation of apoptosis by alternative pre-mRNA splicing. Mol. Cell *19*, 1–13.

Shin, C., and Manley, J.L. (2002). The SR protein SRp38 represses splicing in M phase cells. Cell *111*, 407–417.

Stoilov, P., Lin, C.H., Damoiseaux, R., Nikolic, J., and Black, D.L. (2008). A high-throughput screening strategy identifies cardiotonic steroids as alternative splicing modulators. Proc. Natl. Acad. Sci. USA *105*, 11218–11223.

Taylor, J.K., Zhang, Q.Q., Wyatt, J.R., and Dean, N.M. (1999). Induction of endogenous Bcl-xS through the control of Bcl-x pre-mRNA splicing by antisense oligonucleotides. Nat. Biotechnol. *11*, 1097–1100.

Wahl, M.C., Will, C.L., and Lührmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. Cell *136*, 701–718.

Wang, C., and Youle, R.J. (2009). The role of mitochondria in apoptosis. Annu. Rev. Genet. *43*, 95–118.

Warzecha, C.C., Sato, T.K., Nabet, B., Hogenesch, J.B., and Carstens, R.P. (2009). ESRP1 and ESRP2 are epithelial cell-type-specific regulators of FGFR2 splicing. Mol. Cell *33*, 591–601.

Zhou, A., Ou, A.C., Cho, A., Benz, E.J., Jr., and Huang, S.C. (2008). Novel splicing factor RBM25 modulates Bcl-x pre-mRNA 5′ splice site selection. Mol. Cell. Biol. *28*, 5924–5936.

APPENDIX B: GENOME-WIDE RNAI SCREEN DISCOVERS FUNCTIONAL COUPLING OF ALTERNATIVE SPLICING AND CELL CYCLE CONTROL TO APOPTOSIS REGULATION

Reprinted from
Cell cycle, 2010. **9(22)**: 4419-4421.
with permission from Landes Bioscience

Cell Cycle 9:22, 4419-4421; November 15, 2010; © 2010 Landes Bioscience

Genome-wide RNAi screen discovers functional coupling of alternative splicing and cell cycle control to apoptosis regulation

Qingqing Wang[1] and Pamela A. Silver[1,2,*]

[1]Department of Systems Biology; [2]Wyss Institute for Biologically Inspired Engineering; Harvard Medical School; Boston, MA USA

Alternative splicing (AS) is widely recognized as a major source for biological diversity and regulation. The contribution of AS to eukaryotic development and complexity is likely even greater than previously recognized, as a recent report demonstrates that >90% of human genes are alternatively spliced.[1] The regulation of AS is under the control of cis-elements residing in mRNA sequences and a large repertoire of RNA binding proteins (RNABPs).[2] AS regulation is further integrated with signal transduction and cell-specific cues that control RNABP dynamics and function.[3] While many assays have revealed mechanisms for individual AS regulatory events, global understanding of AS regulation is still at an early phase.[3] Specifically, systematic identification of AS regulators, particularly those upstream of RNABPs, has been difficult via hypothesis-driven studies. Therefore, even in some of the best studied examples of AS it is unclear how AS is connected to other biological processes to carry out complex cellular functions.

We recently completed a whole-genome siRNA screen using a novel reporter strategy to identify factors that regulate AS at all levels. As model AS events, we examined mRNAs for *Bcl-x* and *Mcl1*, Bcl-2 family apoptosis regulators that possess two splice isoforms—anti-apoptotic *Bclx-L* and *Mcl1-L*, and pro-apoptotic *Bclx-S* and *Mcl1-S*.[4] In the reporters, Venus and mCherry fluorescent protein cDNAs were fused to the 3' end of *Bclx* and *Mcl1* genomic sequences, respectively. Premature termination codons (PTC) were then introduced to the alternatively spliced regions exclusive in the long isoforms (**Fig. 1A**). Thus, *Bclx-S* and *Mcl1-S* could be detected as fluorescence fusion proteins, but the PTC prevented expression of fluorescene-tagged long isoforms. A genome-wide siRNA screen using the Bclx reporter revealed 160 factors, that when depleted, led to upregulation of pro-apoptotic *Bclx-S* splicing. These factors were further tested for the Mcl1 reporter, identifying 52 factors that coordinately regulate AS. Surprisingly, knockdown of many cell cycle regulators, including the potent oncogene aurora kinase A (AURKA), dramatically increased *Bclx-S* and *Mcl1-S* isoforms. Further analysis of factors downstream of AURKA linked the splicing changes to direct regulation by the splicing factor ASF/SF2 (SFRS1). Upon AUKRA inhibition, ASF/SF2 was degraded, causing a coordinated shift toward pro-apoptotic splicing in *Bclx* and other apoptosis-related splicing events.

Interestingly, pharmacological inhibition of AURKA by VX-680 (aurora kinase inhibitor) showed that accumulation of pro-apoptotic *Bclx-S* splicing was concomitant to mitotic arrest. However, in actively cycling G_2/M cells, Bclx-S-Venus showed no increase. Also, *Bclx-S* upregulation was not detected during a double-thymidine block that caused S-phase arrest, indicating the specificity of the splicing response to mitotic arrest. Since the current view holds that splicing and transcriptional activity are mostly suppressed in mitosis,[5] we went on to ask if the splicing response preceded mitotic arrest, or whether it could represent a downstream effect of aberrant division in cells that were not yet arrested. Cells were synchronized with double-thymidine, released and treated with VX-680; *Bclx-S* upregulation was found to be comparable to asynchronized cells. Here, the venus-fluorescence was measured 16 hours post thymidine-release, preventing one intervening cell cycle. These data excluded the possibility that the splicing shift was caused by a subset of cells undergoing abnormal cell division before mitotic arrest. Thus, it can be concluded that the *Bclx-S* splicing response preceded mitotic arrest. Consequently, our investigation suggests that pro-death factors accumulate prior to mitotic arrest. This finding provides novel mechanistic clues for how apoptosis occurs in arrested cells, and it is likely that this mechanism extends to other apoptosis regulators.

Aberrant AS is implicated in cancer,[6] but its specific functions are not well understood. In this screen, we found numerous factors that are themselves oncogenes or are frequently misregulated in cancer. Specifically, protein interaction analysis on the top hits showed an AURKA-centered network spanning cell cycle regulators, spliceosome components and tumor suppressors. While aberrant regulation of AURKA is well known to induce spindle checkpoint abrogation, spindle assembly defects and chromosome instability in tumors,[7] our data indicate that AURKA can also contribute to oncogenesis by promoting anti-apoptotic splicing of crucial apoptosis regulators in malignant cells. Conversely, inhibition of AURKA leads to a dramatic elevation in pro-apoptotic splicing isoforms followed by mitotic arrest. Since AURKA is

*Correspondence to: Pamela A. Silver; Email: pamela_silver@hms.harvard.edu
Submitted: 10/25/10; Accepted: 10/26/10
Previously published online: www.landesbioscience.com/journals/cc/article/14051
DOI: 10.4161/cc.9.22.14051
Comment on: Moore MJ, et al. 2010; 142:625–36.

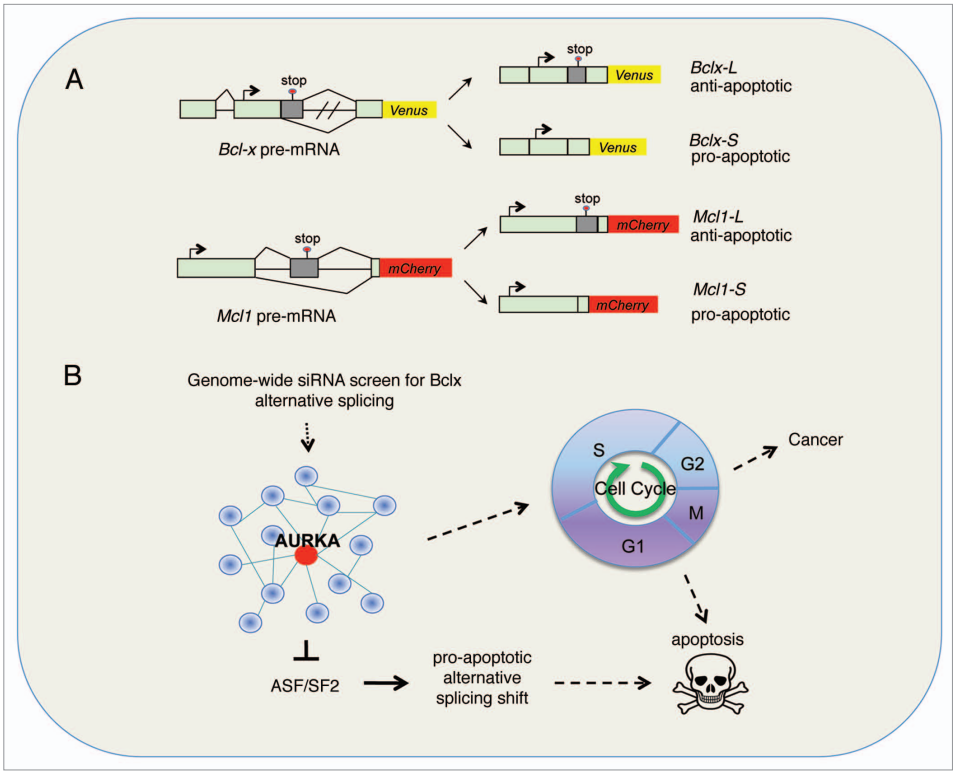

Figure 1. (A) Splicing reporters *Bcl-x* and *Mcl1* are schematized. Green rectangles are exons, with gray areas depicting alternative spliced region. Black lines between rectangles are introns. Arrows denote translational start site. Inserted PTCs exclusive to long isoforms are shown by red "stop signs." (B) A model of the functional coupling of cell cycle control and AS to apoptosis regulation from our genome-wide screen is shown.

a well-known target for anti-tumor therapies, our findings may be useful for new insights in drug development and therapeutic targeting of AURKA.

Taken together, our screen revealed a novel physiological association between the cell cycle and AS. Interestingly, previous studies have hinted at a possible reciprocal link between cell cycle control and AS regulation. For example, many cell cycle regulators have been identified as spliceosome components, including CDC40 and CDC5L, which were strong hits in our study. Moreover, recent large-scale screens identified splicing factors as affecting cell cycle progression and mitotic spindle integrity.[8,9] However, a functional connection between cell cycle and AS has not been emphasized in the literature. Our genome-scale exploration of apoptosis-related AS unexpectedly identified many cell cycle regulators, whose functions spanned multiple aspects of cell cycle control, including regulators of the G_2/M transition, mitotic spindle assembly and microtubule transport. Importantly, downstream of the cell cycle regulators, we discovered an ASF/SF2 mediated 'RNA regulon'* consisting of coordinated *Bclx*, *Mcl1*, *caspase-9* and other apoptosis regulator RNAs. Through this 'RNA regulon', cells respond to upstream signals of cell cycle disruption and carry out a pro-apoptotic AS splicing shift in the RNA members, inducing apoptosis (**Fig. 1B**). Broadly, our screen revealed a novel functional coupling of cell cycle control and AS, where these highly regulated processes are coordinated to regulate apoptosis.

*RNA regulon: 'RNA regulon' hypothesis proposes that groups of functionally related mRNAs are coordinated and regulated by specific, shared RNABPs to quickly react to environmental cues.[10]

Acknowledgements

We thank Michael Moore and Daniel Ducat for critiques of the manuscript. This work was supported by grant NIHGM057476 to P.A.S. from the National Institutes of Health.

References

1. Pan Q, et al. Nat Genet 2008; 40:1413-5.
2. Wahl MC, et al. Cell 2009; 136:701-18.
3. Moore MJ, et al. RNA 2008; 14:197-203.
4. Akgul C, et al. Cell Mol Life Sci 2004; 17:2189-99.
5. Shin C, et al. Cell 2002; 111:407-17.
6. Wang GS, et al. Nat Rev Genet 2007; 8:749-61.
7. Lapenna S, et al. Nat Rev Drug Discov 2009; 8:547-66.
8. Björklund M, et al. Nature 2006; 439:1009-13.
9. Rines DR, et al. Genome Biol 2008; 9:44.
10. Keene JD, et al. Nat Rev Genet 2007; 8:533-43.

APPENDIX C: PQBP1, A FACTOR LINKED TO INTELLECTUAL DISABILITY, AFFECTS ALTERNATIVE SPLICING ASSOCIATED WITH NEURITE OUTGROWTH

Reprinted from
Genes Dev, 2013. **27(6)**: 615-626.
with permission from Cold Spring Harbor Laboratory Press

PQBP1, a factor linked to intellectual disability, affects alternative splicing associated with neurite outgrowth

Qingqing Wang, Michael J. Moore, Guillaume Adelmant, et al.

Genes Dev. 2013 27: 615-626
Access the most recent version at doi:10.1101/gad.212308.112

PQBP1, a factor linked to intellectual disability, affects alternative splicing associated with neurite outgrowth

Qingqing Wang,[1] Michael J. Moore,[2] Guillaume Adelmant,[3,4,5] Jarrod A. Marto,[3,4,5] and Pamela A. Silver[1,6,7]

[1]Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115, USA; [2]Laboratory of Molecular Neuro-Oncology, The Rockefeller University, New York, New York 10065, USA; [3]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA; [4]Blais Proteomics Center, [5]Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA; [6]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, Massachusetts 02115, USA

Polyglutamine-binding protein 1 (PQBP1) is a highly conserved protein associated with neurodegenerative disorders. Here, we identify PQBP1 as an alternative messenger RNA (mRNA) splicing (AS) effector capable of influencing splicing of multiple mRNA targets. PQBP1 is associated with many splicing factors, including the key U2 small nuclear ribonucleoprotein (snRNP) component SF3B1 (subunit 1 of the splicing factor 3B [SF3B] protein complex). Loss of functional PQBP1 reduced SF3B1 substrate mRNA association and led to significant changes in AS patterns. Depletion of PQBP1 in primary mouse neurons reduced dendritic outgrowth and altered AS of mRNAs enriched for functions in neuron projection development. Disease-linked PQBP1 mutants were deficient in splicing factor associations and could not complement neurite outgrowth defects. Our results indicate that PQBP1 can affect the AS of multiple mRNAs and indicate specific affected targets whose splice site determination may contribute to the disease phenotype in PQBP1-linked neurological disorders.

[*Keywords*: PQBP1; alternative splicing; neurite outgrowth; neurological disorders]

Supplemental material is available for this article.

Received December 14, 2012; revised version accepted February 21, 2013.

Alternative splicing (AS) is a major source for biological diversity and a crucial determinant of cell identity and fate. To date, >90% of human precursor messenger RNAs (pre-mRNAs) are reported to undergo AS, with immense variation across tissue types and developmental stages (Wang et al. 2008). The regulation of AS is carried out by the interaction of RNA-binding proteins (RBPs), notably hnRNPs and SR proteins, with *cis*-elements in the pre-mRNAs (Black 2003). The dynamics of RBP *cis*-element recognition are further modulated by a wide spectrum of spliceosome-associated proteins and are responsive to signal transduction networks (Varani and Nagai 1998; Wahl et al. 2009). Tissue-specific AS regulation is especially prominent in the nervous system (Castle et al. 2008), where neurons must promptly respond to complex stimuli and perform downstream functions with flexibility and precision (Dredge et al. 2001). Accordingly, neuron-specific AS networks modulate such diverse functions, including neurotransmission, receptor activity, and ion channel function, and thus impact higher brain functions, including cognition, coordination, and learning (Grabowski and Black 2001).

Aberrant AS is a primary cause for many human diseases (Wang and Cooper 2007). Diverse neurological disorders are closely linked to aberrant splicing and can be classified as *cis*-acting and *trans*-acting splicing defects (Licatalosi and Darnell 2006). *Cis*-acting defects originate from mutations in the disease-associated gene that change sequences important for the correct splicing of the pre-mRNA. A representative example is frontotemporal dementia with Parkinsonism linked to chromosome 17, which arises from mutations causing missplicing of the mRNA for the microtubule-associated protein tau (Licatalosi and Darnell 2006). *Trans*-acting defects arise from mutations or misregulation of splicing regulatory factors. One example is retinitis pigmentosa, characterized by a progressive loss of rod photoreceptor cells, which results from mutations in HPRP3, PRP31, and PRPC8—factors crucial for the assembly and functioning of U small nuclear ribonucleoproteins (snRNPs) in pre-mRNA splicing (McKie et al. 2001; Vithana et al. 2001; Chakarova et al. 2002). Although increasing numbers of

[7]**Corresponding author**
Email pamela_silver@hms.harvard.edu
Article is online at http://www.genesdev.org/cgi/doi/10.1101/gad.212308.112.

splicing-related disorders have been identified, the molecular link between AS regulation and the disease phenotype remains to be elucidated.

Polyglutamine-binding protein 1 (PQBP1) is a highly conserved protein with mutations associated with ~10 familial X-linked mental retardation (XLMR) diseases (Kalscheuer et al. 2003; Stevenson et al. 2005). Patients with these diseases share similar symptoms, such as intellectual deficiency, microcephaly, and short stature (Kalscheuer et al. 2003; Stevenson et al. 2005). Although these diseases were separately discovered and originally thought to be unrelated, they were recently grouped together as "Renpenning syndrome" because of the common symptomatic manifestation and mutations in PQBP1 (Stevenson et al. 2005).

The molecular function of PQBP1 remains unclear. PQBP1 is linked to transcription regulation based on its interaction with RNA polymerase II and other transcription factors (Waragai et al. 1999; Okazawa et al. 2002). Recent studies also reported that PQBP1 was dynamically associated with stress granules and interacted with dynactin components, indicating a possible role for PQBP1 in RNA transport or cytoplasmic RNA processing (Kunde et al. 2011). Also, PQBP1 has been implicated in pre-mRNA splicing: PQBP1 interacts with known splicing factors such as WBP11 and U5-15KD (Komuro et al. 1999; Waragai et al. 2000) and is part of the early spliceosome (Makarova et al. 2004). Consistent with this view, a disease-associated mutant of PQBP1 decreased splicing efficiency (Tapia et al. 2010).

We previously identified PQBP1 in a screen for factors that affect the AS of the mRNA for Bcl-X, a mammalian apoptosis regulator (Moore et al. 2010). Here, we show that PQBP1 affects the AS of numerous other pre-mRNAs in human cells. In addition, we used whole-transcriptome sequencing to identify a global set of PQBP1 AS targets in primary mouse neurons, which revealed that PQBP1 coordinates important neuronal functions and maintains normal neurite development. Collectively, our findings indicate that aberrant AS contributes to the disease pathology of PQBP1-associated neurological disorders.

Results

PQBP1 affects the AS of mRNAs encoding apoptotic factors

We previously identified PQBP1 as a potential effector for *Bcl-x* splicing in HeLa cells in a genome-wide siRNA screen using a fluorescence-based minigene assay (Fig. 1A; Moore et al. 2010). To verify the effect of PQBP1 on the AS of *Bcl-x*, we examined the splicing of endogenous *Bcl-x* after depletion of PQBP1 with two different siRNAs (Fig. 1B) using RT–PCR primers flanking the alternatively spliced region (Fig. 1A). As expected, depletion of PQBP1 favored a shift from the anti-apoptotic isoform *Bcl-xL* toward the proapoptotic isoform *Bcl-xS* via increased use of an alternative 5′ splice site in the second exon (Fig. 1A,C; Supplemental Fig. S1A,B). These results confirmed that PQBP1 affects the AS of *Bcl-x* in HeLa cells.

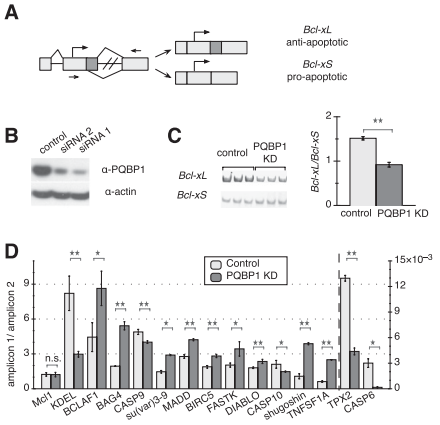

Figure 1. AS targets of PQBP1 in HeLa cells. (*A*) Schematic of *Bcl-x*. Light rectangles are exons, with dark color marking the alternatively spliced region, and black lines are introns. Arrows show translation start sites, and arrowheads mark quantitative PCR primer sets. (*B*) Western blot of HeLa cell lysates where PQBP1 is knocked down by siRNA. The control sample was treated with a siRNA targeting firefly luciferase mRNA. Antibodies are listed on the *right*. Knockdown (KD) with two different siRNAs is shown. (*C*) Gel analysis of RT–PCR of *Bcl-x* splicing isoforms for control and PQBP1 knockdown samples is shown (*left* panel) with quantification (*right* panel). Mean of three independent measurements ± SD are shown. Changes were tested by one-way ANOVA. (**) Statistically significant with *P*-value < 0.01. See also Supplemental Figure S1, A and B. (*D*) Quantification of real-time PCR data on 14 mRNAs that showed significant AS changes between control and PQBP1 knockdown samples, with *Mcl1* (not a target of PQBP1) as the negative control. The *Y*-axis depicts the ratio between amplicons for constitutive and alternative exonic regions; *TPX-2* and *CASP6* follow the *Y*-axis on the *right* side, while the rest follow the *Y*-axis on the *left* side. Light columns show the ratio of two amplicons from the control sample, and dark columns show the ratio of two amplicons from the PQBP1 knockdown sample. Mean of three independent measurements ± SD are shown. The difference of the ratio between control and PQBP1 knockdown samples was analyzed by one-way ANOVA test. (*) Statistically significant with *P*-value < 0.03; (**) statistically significant with *P*-value < 0.01; (n.s.) not significant. See also Supplemental Figure S1C.

We now show that PQBP1 affects additional AS events. We analyzed 31 mRNAs encoding apoptosis-related factors with annotated alternatively spliced isoforms. We designed primers to amplify the alternatively spliced exonic region and constitutive exonic region of each mRNA, and the ratio of the two amplicons was compared between the control and PQBP1 knockdown samples by RT–PCR. Approximately half of these mRNAs showed significant changes in AS when PQBP1 was knocked down (Fig. 1D; Supplemental Fig. S1C). These results provided further evidence that PQBP1 affects the AS of many mRNAs.

PQBP1 is associated with splicing factors

To further clarify PQBP1's role in AS regulation, we systematically determined the spectrum of proteins that are associated with PQBP1. In addition, to understand the aberrant function of disease-related PQBP1 mutants, we compared the associated protein profile of wild-type PQBP1 and two mutant variants identified in Renpenning syndrome patients.

PQBP1 protein contains a WW domain, a polar amino acid-rich domain, and a C-terminal domain with a nuclear localization signal (NLS) (Fig. 2A; Waragai et al. 1999). The mutations in PQBP1 identified from Renpenning syndrome patients can be classified into two categories: frameshifting mutations in the AG dinucleotide hexamer in the polar amino acid-rich domain and missense mutations in the WW domain (Musante et al. 2010). Here, we examined two representative disease-linked variants: one with an AG dinucleotide deletion (ΔAG) in the polar amino acid-rich domain, resulting in a truncated protein, and one with a point mutation (Y65C) in the WW domain (Fig. 2A; Kalscheuer et al. 2003). PQBP1 localizes predominantly in the nucleus (Fig. 2B; Waragai et al. 1999; Komuro et al. 1999; Kalscheuer et al. 2003). Interestingly, both ΔAG and Y65C mutants localize to the nucleus (Fig. 2B), although ΔAG loses a potential NLS (Fig. 2A), indicating that there must be an additional unidentified NLS or that it enters together with another protein.

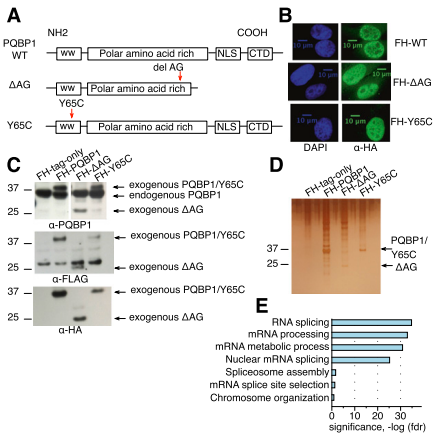

Figure 2. The spectrum of proteins associated with PQBP1 and disease-linked mutants. (*A*) Schematics of the protein domain organization of wild-type (WT) PQBP1 and two disease-linked mutants. Boxes are protein domains, with names shown in the center. (WW) WW domain; (CTD) C-terminal domain. Red arrowheads show the mutation sites in the disease-linked PQBP1 variants. (*B*) Immunofluorescence staining of HeLa cells stably expressing Flag-HA-tagged wild-type PQBP1, ΔAG, or Y65C. Cells were fixed and stained with antibodies or dyes listed at the *bottom* of each panel column. (*C*) Western blot of HeLa cell lines stably expressing empty Flag-HA vector, Flag-HA-tagged wild-type PQBP1, Flag-HA-tagged ΔAG, and Flag-HA-tagged Y65C. The antibodies used are indicated at the *bottom* of each panel. (*D*) Silver staining of proteins copurified with Flag-HA-tagged empty plasmid, Flag-HA-tagged wild-type PQBP1, Flag-HA-tagged ΔAG, or Flag-HA-tagged Y65C. (*E*) GO term enrichments of proteins that were pulled down with wild-type PQBP1. Statistics were calculated with human genes as the background. The *X*-axis is the corrected *P*-value (false discovery rate [FDR]) in negative log for enrichment.

Proteins associated with wild-type PQBP1 were strongly enriched for known or potential splicing factors and spliceosome components (Fig. 2E; Supplemental Table S1). HeLa cells expressing stably integrated Flag-HA-tagged full-length PQBP1, ΔAG, or Y65C were generated by stable integration of retroviral constructs and shown by Western blotting to express exogenous proteins at a level comparable with endogenous PQBP1 (Fig. 2C). Tandem affinity purification with Flag and HA antibodies followed by mass spectroscopy was performed to identify PQBP1- or mutant-associated proteins (Fig. 2D; Supplemental Table S1). WBP11, a previously known PQBP1 interactor (Komuro et al. 1999), was detected in all three replicate purifications of wild-type PQBP1, confirming the validity of our approach (Supplemental Table S1). U5-15KD, another proposed interactor of PQBP1 (Waragai et al. 2000), was not identified, possibly because of cell specificity or because the interaction is not strong enough to be detected in our tandem affinity purification. ΔAG showed significantly reduced splicing factor association compared with wild-type PQBP1, while Y65C lost associations with all splicing-related factors (Supplemental Table S1). These results established an association, either direct or indirect, of PQBP1 to the splicing machinery and identified the WW domain as important for linking PQBP1 to splicing complexes. This finding is consistent with the idea that disease-linked mutations may disrupt splicing-related functions of PQBP1.

PQBP1 influences SF3B1's (subunit 1 of the splicing factor 3B [SF3B] protein complex) recognition of splicing sites

To explore the mechanism of PQBP1 in AS regulation, we first examined whether PQBP1 directly binds pre-mRNAs whose splicing it affects. Although previous studies showed PQBP1 bound to polyrG resin in vitro (Komuro et al. 1999), PQBP1 protein itself has no known RNA-binding domains (Fig. 2A; Lunde et al. 2007). We used the CLIP (cross-linking and immunoprecipitation) method to determine whether PQBP1 directly binds its targets in vivo (Ule et al. 2003). Live HeLa cells were exposed to UV light to cross-link RNA–protein complexes, and then PQBP1 was immunopurified from cell lysates (Fig. 3A). RT–PCR quantification of *Bcl-x* and nine other target mRNAs (Fig. 1D) showed no enrichment in PQBP1 immunoprecipitations compared with an irrelevant immunoglobulin G (IgG) control. These results suggest that PQBP1 does not bind AS target mRNAs directly, with the caveat that it may fail to be effectively cross-linked under the

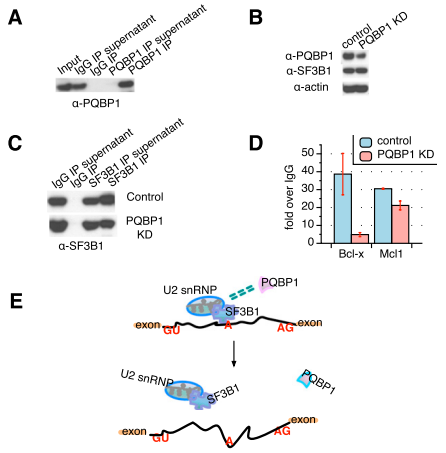

Figure 3. SF3B1 function is affected by PQBP1. (*A*) Western blot of PQBP1 CLIP together with an irrelevant IgG control, probed with PQBP1 antibody. PQBP1 was not detected in depleted supernatants or the IgG pull-down, indicating efficient and specific antibody binding of PQBP1. (*B*) Western blot of SF3B1 CLIP from PQBP1 knockdown (KD) samples and samples treated with siRNA targeting firefly luciferase mRNA as control. Shown are total cell lysates of control (lane *1*) and PQBP1 knockdown (lane *2*) HeLa cells probed with antibodies indicated at the *left*. (*C*) Western blot of SF3B1 CLIP from PQBP1 knockdown samples and samples treated with siRNA targeting firefly luciferase mRNA as control. Lanes *1* and *2* are the IgG controls, and lanes *3* and *4* are SF3B1 CLIP stained by SF3B1 antibody for control and PQBP1 knockdown samples. (*D*) Quantification of *Bcl-x* and *Mcl1* mRNA enrichment in the SF3B1–RNA complex as from SF3B1 CLIP in control and PQBP1 knockdown samples, respectively, by RT–PCR. Data values normalized to the IgG control and mean of three independent measurements ± SD are shown. (*E*) Proposed model mechanism of PQBP1 AS regulation. A schematic of part of a hypothetical mRNA is shown. Orange rectangles are exons, and the black line is the intron. "GU" marks the 5′ splice site, "A" marks the branch site recognized by the U2 snRNP with help from components like SF3B1, and "AG" marks the 3′ splice site. Green dashed double lines indicate the association between PQBP1 and SF3B1. Through an association with SF3B1 and other splicing factors, PQBP1 influences the recruitment of U2 snRNP to specific target sites, thus affecting splicing decisions on a subset of mRNAs. Loss or mutations of PQBP1 interrupt the splicing factor association and the recognition of specific splice sites. See the Discussion.

conditions where other known splicing factors are observed to bind.

PQBP1, on the other hand, affects the association of the spliceosome factor SF3B1 with a target pre-mRNA. Among the proteins that are associated with PQBP1 are SF3B protein complex members (Supplemental Table S1), which form part of the U2 snRNP. The function of the SF3B complex is critical for the assembly of the spliceosome complex A as well as the successful execution of the splicing reaction (Wahl et al. 2009). Furthermore, SF3B1 was reported to regulate the AS of *Bcl-x* through direct interaction with a *cis*-element in the *Bcl-x* mRNA (Massiello et al. 2006), and we observed that CLIP with SF3B1 showed a 40-fold enrichment of *Bcl-x* binding over an IgG control (Fig. 3D, blue column over *Bcl-x*).

To test whether the recognition of *Bcl-x* by SF3B1 is influenced by PQBP1, we knocked down PQBP1 in HeLa cells and reanalyzed the enrichment of *Bcl-x* in the SF3B1–RNA complex (Fig. 3B,C). PQBP1 depletion led to an eightfold decrease of *Bcl-x* mRNA levels in SF3B1 CLIP compared with a control knockdown (Fig. 3D), although the level of SF3B1 protein remained unchanged (Fig. 3B), and the steady-state level of *Bcl-x* mRNA did not change upon PQBP1 knockdown (Supplemental Fig. S2).

To further test whether PQBP1's impact on SF3B1–RNA recognition is specific to the targets of PQBP1, we analyzed the mRNA of *Mcl1* in the SF3B1–RNA complex. Like Bcl-X, Mcl1 is a mammalian apoptosis regulator and has two splicing isoforms: anti-apoptotic *Mcl1-L* and proapoptotic *Mcl1-S* (Akgul et al. 2004). However, while knockdown of SF3B1 significantly shifted the splicing of *Mcl1* toward *Mcl1-S* (Moore et al. 2010), PQBP1 knockdown had little effect (Fig. 1D; Moore et al. 2010). We therefore hypothesized that the knockdown of PQBP1 should have relatively minor influence on *Mcl1* mRNA enrichment in the SF3B1–RNA complex. Indeed, CLIP data showed only a 1.4-fold decrease of *Mcl1* mRNA enrichment in the SF3B1–RNA complex in the PQBP1 knockdown sample compared with the control (Fig. 3D). Additionally, the decrease was mostly likely due to a *Mcl1* mRNA steady-state level change of 1.5-fold (Supplemental Fig. S2). Therefore, PQBP1 influenced the recognition of SF3B1 to one mRNA whose AS is affected by PQBP1 but not for an mRNA where AS was regulated by SF3B1 but not PQBP1. These results do not rule out possibilities that PQBP1 may affect mRNA AS through other interactions or that other splicing factors may assist in the process.

Loss of PQBP1 causes defects in neurite outgrowth

Previous work found that PQBP1 is most highly expressed in the CNS of embryonic or newborn rodents, with the peak around birth (Qi et al. 2005). Considering the tissue-specific expression pattern of PQBP1 as well as its disease association, we examined the role of PQBP1 in AS modulation in embryonic rodent neurons.

PQBP1 is highly expressed in embryonic mouse cortical and hippocampal neurons, as shown by Western blots (Fig. 4A). In addition, PQBP1 localizes predominantly in the nucleus of neurons (Fig. 4B, top panels), as shown previously in other cell types (Waragai et al. 1999; Komuro et al. 1999; Kalscheuer et al. 2003). A closer examination demonstrated that PQBP1 colocalizes with SC35—a splicing factor in nuclear speckles, a nuclear domain enriched for splicing regulatory factors (Fig. 4B, bottom panels; Supplemental Fig. S3A; Spector and Lamond 2011).

To understand the importance of PQBP1 in neuron development, we knocked down PQBP1 in mouse embryonic primary cortical neurons with shRNAs delivered by lentivirus (Fig. 4C). Neurons were first examined by immunostaining of the microtubule-associated protein 2 (MAP2), a marker for dendrites. Interestingly, upon PQBP1 knockdown, the number of projections and length of neurites were decreased, and the connection between neurons degenerated considerably (Fig. 4D). A second shRNA against PQBP1 led to a similar defect in neurite morphology (Supplemental Fig. S3B,C). To determine whether this defect is simply due to cell apoptosis, we compared the number of DAPI-stained intact nuclei after fixation as well as the percentage of Annexin V-stained cells between control and PQBP1 knockdown neurons (Fig. 4E,F). No significant differences were observed, indicating that the defects observed from PQBP1 knockdown are not from neuron apoptosis.

PQBP1 knockdown neurons showed substantially diminished dendrite branching and dendrite length (Fig. 4G,H). To quantify the changes in neurite outgrowth, we transiently transfected neurons with a nontargeting control shRNA or a PQBP1 targeting shRNA together with a GFP vector (Fig. 4G). Sholl analysis was then performed on neurons that were GFP-highlighted for control and PQBP1 knockdown (Fig. 4H; Sholl 1953).

To confirm that the dendritic outgrowth defect observed was a direct effect of PQBP1 knockdown, we examined whether the defect could be rescued by re-expressing wild-type PQBP1 in comparison with the disease-associated PQBP1 mutants. shRNA-resistant, Flag-HA-tagged cDNAs encoding human PQBP1, ΔAG, and Y65C were coexpressed in mouse cortical neurons with shRNAs targeting endogenous mouse PQBP1 (Fig. 5A; Supplemental Fig. S3D). Western blots confirmed the knockdown of endogenous mouse PQBP1 and the concomitant expression of exogenous dual-tagged human PQBP1 and variants (Fig. 5B). Total PQBP1 expression in all three rescue samples was compatible with the level of PQBP1 in the control sample, with the exception of ΔAG, which reached only 68% of the control (Fig. 5B). To quantitatively compare neurite outgrowth between mutant PQBP1 and wild-type rescue, we transfected mouse cortical neurons with the rescue constructs (Fig. 5C) and performed Sholl analysis after 5 d. A significant restoration of dendrite density and length was observed in wild-type PQBP1-rescued neurons, while ΔAG and Y65C showed less improvement (Fig. 5D,E,F). In the case of ΔAG, it is possible that insufficient expression of ΔAG may contribute to the deficiency of rescue.

Combining the results above, we conclude that PQBP1 is important in maintaining neurite projection and out-

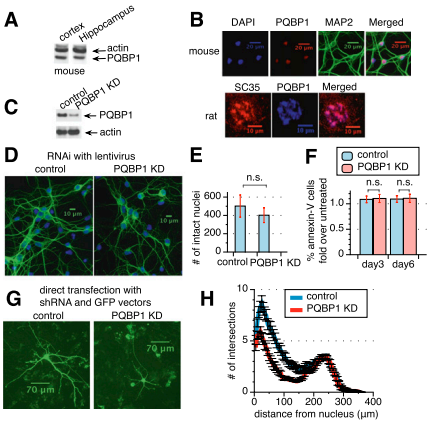

Figure 4. PQBP1 affects dendritic outgrowth and branching in mouse cortical neurons. (*A*) Western blot of PQBP1 expression in mouse embryonic cortical and hippocampal neurons, compared with endogenous β-actin expression. (*B*) Immunofluorescence staining of mouse embryonic primary cortical neurons (the four *top* panels) and rat embryonic primary cortical neurons (the three *bottom* panels). Neurons were fixed and stained with antibodies or dyes listed at the *top* of each panel. See also Supplemental Figure S3A. (*C*) Western blot of mouse embryonic cortical neurons infected with virus from a nontargeting control shRNA or a PQBP1 targeting shRNA. See also Supplemental Figure S3B. (*D*) Immunofluorescence staining of mouse primary embryonic cortical neurons 5 d after infection with virus from a nontargeting control shRNA or a PQBP1 targeting shRNA. (Green) MAP2 staining; (blue) DAPI staining. See also Supplemental Figure S3C. (*E*) Comparison of the number of intact nuclei between neuron samples 5 d after infection with virus from a nontargeting control shRNA or a PQBP1 targeting shRNA. Neurons were fixed and stained with DAPI. Healthy and intact nuclei were counted for each sample. Mean of eight independent measurements ± SD is shown. The difference in the number of intact nuclei between control and PQBP1 knockdown samples was analyzed by one-way ANOVA test. (n.s.) Not significant. (*F*) Comparison of the percentage of Annexin V-stained neurons between samples infected with virus from a nontargeting control shRNA or a PQBP1 targeting shRNA. The number of neurons with Annexin V staining and the percentage of Annexin V-positive neurons were calculated and normalized to neuron samples that were not treated by virus. The *Y*-axis depicts the fold of the percentage of Annexin V-positive neurons in control or PQBP1 knockdown samples to samples that were untreated. Neurons on day 3 and day 6 after infection/plating were studied. The mean of six independent measurements ± SD is shown. The difference in the percentage of Annexin V-positive neurons between control and PQBP1 knockdown samples was analyzed by one-way ANOVA test. (n.s.) Not significant. (*G*) Immunofluorescence of individual neurons that were transiently transfected with either a nontargeting control shRNA or a PQBP1 targeting shRNA together with a GFP vector. Neurons were grown on a layer of feeder glial cells and examined 5 d after transfection. (*H*) Sholl analysis of dendritic branching for neurons transiently transfected with a nontargeting control shRNA or PQBP1 targeting shRNA together with a GFP vector. Concentric circles were drawn at intervals of 5 μm. The number of dendritic branch intersections with each concentric circle was counted (mean ± SE, $n = 16–17$). The *X*-axis depicts the distance of the concentric circle from the neuron soma center, and the *Y*-axis depicts the number of intersections the corresponding circle encounters with dendrites.

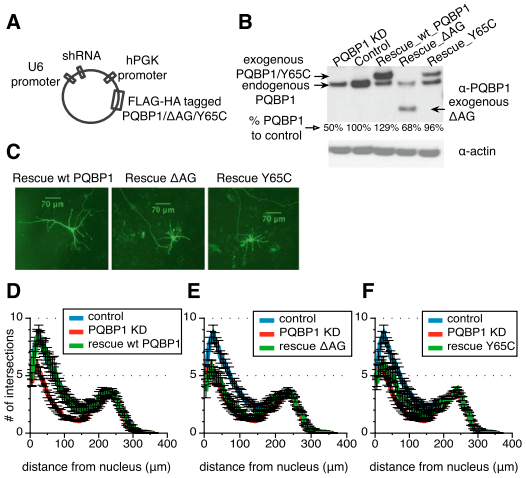

Figure 5. Wild-type (WT) PQBP1 but not the mutants rescues the neuron morphological defects. (*A*) Schematic of the pLKO.1-derived lentiviral constructs used to generate viruses for the rescue experiments. See also Supplemental Figure S3D. (*B*) Western blot of neuron samples that were infected with virus from an unmodified PQBP1 targeting shRNA, a nontargeting control shRNA, or PQBP1 targeting shRNA constructs coexpressing exogenous wild-type PQBP1, ΔAG, or Y65C. Antibodies used are listed at the *right*. Quantification of total PQBP1 level for each sample compared with control is shown *below* each lane. (*C*) Immunofluorescence of individual neurons that were transiently transfected with PQBP1 targeting shRNA constructs coexpressing exogenous wild-type PQBP1, ΔAG, or Y65C together with a GFP vector. Neurons were grown on a layer of feeder glial cells. (*D–F*) Sholl analysis of dendritic branching for neurons that were transiently transfected with a nontargeting control shRNA, an unmodified PQBP1 targeting shRNA, or PQBP1 targeting shRNA constructs coexpressing wild-type PQBP1 (*D*), ΔAG (*E*), or Y65C (*F*) (mean ± SE; n = 10–17).

growth in neuron development. Moreover, deviation in neurite structure may be part of the defects brought about by the disease-associated mutants of PQBP1.

Coupling of neuron projection and PQBP1 function

Dendrite structure is the crucial basis for neuron network formation and information processing. To clarify the molecular function of PQBP1 in maintaining neurite outgrowth, we systematically profiled the AS targets of PQBP1 in mouse embryonic cortical neurons by RNA sequencing (RNA-seq).

We considered RNA-seq reads that were directly mapped over splice junctions as the most straightforward evidence to quantify the splicing changes. For each splice junction, a 5′ initiation site (5′ IS) is defined as the final nucleotide position of the upstream exon, and the 3′ end site (3′ ES) is defined as the first nucleotide position of the downstream exon (Fig. 6A). An AS event is then defined as a set of junctions that share the same 5′ IS or the same 3′ ES, with each junction in the set referred to as a sub-AS junction (Fig. 6A). This definition covers all six primary AS patterns except for intron retention, which was not considered in our analysis (Supplemental Fig. S4A; Black 2003). An AS event under an experimental condition is represented by the distribution of read counts mapped to each sub-AS junction within the AS event; i.e., the "usage" of each sub-AS junction under that condition. We developed methods to identify AS events that experienced significant differential "usage" of any sub-AS junction between control and PQBP1 knockdown conditions (for details, see the Supplemental Material) and defined these AS events as AS targets of PQBP1. In total, 457 targets were identified in the process (Supplemental Fig. S4B; Supplemental Table S2).

To validate PQBP1 targets identified by RNA-seq, we picked 10 candidate targets together with two candidate nontargets from the analysis and designed primers to amplify each splicing isoform of each gene. We then compared the ratio of the isoform expression levels by RT–PCR for control and PQBP1 knockdown samples. All 10 identified targets exhibited significant AS shift upon PQBP1 knockdown, while the two nontargets showed no discernible changes (Fig. 6B). This result demonstrated the sensitivity and specificity of our computational method to identify AS targets of PQBP1.

The PQBP1 AS targets showed tissue-specific functional enrichment, with strong gene ontology (GO) enrichments for neuron projection development/morphogenesis, dendrite development, and axonogenesis (Fig. 6C; Supplemental Table S4) versus genes that were expressed in mouse embryonic cortical neurons profiled from the RNA-seq data. This observation was in accordance with the neurite outgrowth defect observed when PQBP1 was knocked down (Fig. 4D,G,H). Targets of PQBP1 also showed strong enrichment in RNA splicing, synaptic transmission, chromatin modification, cell–cell signaling, phosphate metabolic process, neurotransmitter transport, and ADP-ribosylation factor (ARF) protein signal transduction (Fig. 6C; Supplemental Table S4). The GO term enrichment profile for PQBP1's AS targets is specific compared with high-throughput studies on other proteins in neurons, especially RNA splicing, chromatin modification, dendrite development-related functions, and ARF signal transduction (Ule et al. 2005; Gehman et al. 2011; Charizanis et al. 2012). *Bcl-x* AS is not conserved from humans to mice and was not identified as a target in this analysis. Among the 14 AS targets identified in HeLa cells that were related to apoptosis control (Fig. 1D), only the mRNA of *BCLAF1* was recognized as an AS target of

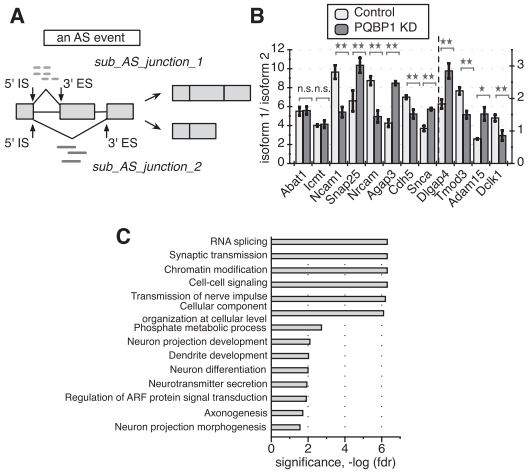

Figure 6. PQBP1's AS targets in mouse embryonic cortical neurons. (*A*) A schematic of a defined AS event and quantification of splicing changes. A hypothetical alternatively spliced mRNA is shown, with light rectangles representing exons and black lines between exons representing introns. The middle exon is the alternatively spliced exon. The 5′ IS is the start of a splicing junction, and the 3′ ES is the end of the junction. The AS event shown here includes two sub-AS junctions that share the same 5′ IS. Light dashed short lines mark RNA-seq reads that are mapped to sub-AS junction 1, and dark solid short lines mark reads mapped to sub-AS junction 2. The distribution of reads mapped to the two sub-AS junctions was used for AS quantification in our analysis. See also Supplemental Figure S4, Supplemental Table S2, and the Supplemental Material. (*B*) RT–PCR validation of 10 randomly picked AS events identified as PQBP1 targets and two randomly picked AS events identified as non-PQBP1 targets in the computational analysis. *Abat1* and *Icmt* were predicted non-PQBP1 targets. The *Y*-axis depicts the ratio between isoform 1 and isoform 2. *Dlgap4*, *Tmod3*, *Adam15*, and *Dclk1* follow the *Y*-axis on the *right* side, while the rest follow the *Y*-axis on the *left* side. The mean of three independent measurements ± SD is shown. The differences between control and PQBP1 knockdown (KD) samples were analyzed by one-way ANOVA test. (*) Statistically significant with *P*-value < 0.05; (**) statistically significant with *P*-value < 0.01; (n.s.) not significant. (*C*) GO term functional enrichment of AS targets of PQBP1. Statistics were calculated with all genes that are expressed in mouse embryonic cortical neurons (profiled from RNA-seq data) as the background. The GO term enrichment profile here is specific compared with RNA-seq studies for other proteins in neurons (see the text). The *X*-axis is the corrected *P*-value (FDR) in negative log for enrichment. See also Supplemental Table S4.

PQBP1 in mouse neurons, indicating the tissue specificity of PQBP1's regulation in AS.

Considering the physical and potential functional association of PQBP1 and SF3B1, we also knocked down SF3B1 in mouse cortical neurons and examined the AS of the 10 verified targets of PQBP1 (Supplemental Fig. S5A). Interestingly, all 10 targets also showed a significant AS shift upon SF3B1 knockdown (Supplemental Fig. S5B).

Among the top hits of PQBP1 AS targets was NCAM-140, a major isoform of neural cell adhesion molecules 1 (Ncam1) (Fig. 6B). Specifically, a 30-nucleotide (nt) exon between exons 7 and 8 is alternatively spliced in NCAM-140 and is designated VASE (variable alternatively spliced exon) (Fig. 7A). The VASE-excluding isoform of NCAM-140 ($VASE^-$) promotes neurite outgrowth, while the VASE-including isoform ($VASE^+$) inhibits neurite outgrowth (Doherty et al. 1992). The embryonic brain mainly expresses $VASE^-$ with gradual ascendancy until final dominance of $VASE^+$ expression toward adulthood, except in regions such as the olfactory bulb, where reformation of synapses and turnover of neurons continue (Small and Akeson 1990). Thus, the AS of VASE in NCAM-140 is a critical switch for neuronal plasticity. To confirm PQBP1's effect on VASE, we designed primers flanking the VASE exon and examined $VASE^+$ and $VASE^-$ isoforms by RT–PCR and electrophoresis (Fig. 7A). $VASE^+$ was increased upon PQBP1 knockdown, while $VASE^-$ decreased (Fig. 7B), resulting in a significant decrease in the ratio of $VASE^-/VASE^+$ expression in response to PQBP1 shRNA versus a control hairpin (Fig. 6B, *Ncam1*). These results were confirmed with a second shRNA targeting PQBP1 (Supplemental Fig. S6).

As a further confirmation that PQBP1 impacts the AS of NCAM-140, we examined AS of NCAM-140 after PQBP1 knockdown in neurons coexpressing the wild-type PQBP1, ΔAG, and Y65C constructs from Figure 5A. Interestingly, neurons rescued with wild-type PQBP1 restored the ratio of $VASE^-/VASE^+$ to levels observed in control samples (Fig. 7C). In contrast, ΔAG and Y65C were less effective in correcting the $VASE^-/VASE^+$ ratio (Fig. 7C).

To test whether NCAM-140 is one of the PQBP1 targets responsible for the defects observed in cells lacking PQBP1, we examined the effect of restoring the ratio of $VASE^-$ and $VASE^+$. We coexpressed $VASE^-$ under the hPGK promoter in the lentiviral shRNA vector as described above (Fig. 5A). Knockdown of PQBP1 was verified by Western blot (Fig. 7D), and the change in the ratio of $VASE^-/VASE^+$ was verified by RT–PCR (Fig. 7E). Neurons were transfected with the rescue construct (Fig. 7F), and Sholl analysis on transfected neurons displayed a significant recovery from the neurite outgrowth defect caused by PQBP1 knockdown (Fig. 7G). These results confirm the role of PQBP1 in the AS of NCAM-140 and modulating neurite outgrowth.

Discussion

In this study, we showed that the disease-associated factor PQBP1 plays a role in the AS of mRNAs. We

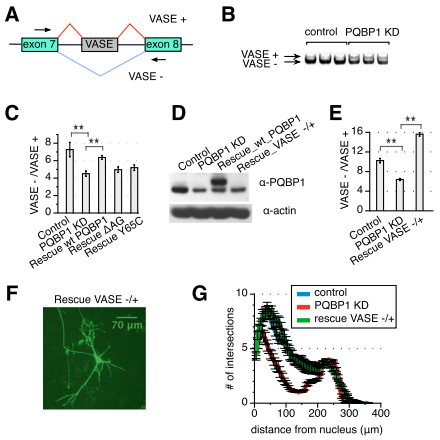

Figure 7. PQBP1-modulated AS of NCAM-140 influences neurite outgrowth. (*A*) A schematic of alternatively spliced region of NCAM-140. Green rectangles are exons, and black lines are introns. The gray rectangle marks the VASE. Arrowheads are primers designed for RT–PCR and electrophoresis. (*B*) NCAM-140 VASE splicing isoforms for neuron samples infected with virus from a nontargeting control shRNA or a PQBP1 targeting shRNA were analyzed by RT–PCR and gel electrophoresis. See also Supplemental Figure S6. (*C*) RT–PCR quantification for $VASE^-$ and $VASE^+$. The ratio of the two isoform levels are compared among neurons infected with virus from a nontargeting control shRNA, an unmodified PQBP1 targeting shRNA, or PQBP1 targeting shRNA constructs coexpressing exogenous wild-type (WT) PQBP1, ΔAG, or Y65C. The mean of three independent measurements ± SD is shown. The differences between samples were analyzed by one-way ANOVA test. (**) Statistically significant with *P*-value < 0.01. (*D*) Western blot of neurons infected with virus from a nontargeting shRNA, an unmodified PQBP1 targeting shRNA, a PQBP1 targeting shRNA construct coexpressing exogenous wild-type PQBP1, or a PQBP1 targeting shRNA construct coexpressing exogenous $VASE^-$ to restore the ratio of $VASE^-/VASE^+$. (*E*) RT–PCR quantification for $VASE^-$ and $VASE^+$ expression levels of neurons infected with virus from a nontargeting shRNA, an unmodified PQBP1 targeting shRNA, or a PQBP1 targeting shRNA construct coexpressing exogenous $VASE^-$ to restore the ratio of $VASE^-/VASE^+$. The mean of three independent measurements ± SD is shown. The differences between samples were analyzed by one-way ANOVA test. (**) Statistically significant with *P*-value < 0.01. (*F*) Immunofluorescence of individual neurons that were transiently transfected with a PQBP1 targeting shRNA construct coexpressing exogenous $VASE^-$ to restore the ratio of $VASE^-/VASE^+$ together with a GFP vector. Neurons were grown on a layer of feeder glial cells. (*G*) Sholl analysis of dendritic branching for individual neurons that were transiently transfected with a nontargeting control shRNA, an unmodified PQBP1 targeting shRNA, or a PQBP1 targeting shRNA construct coexpressing exogenous $VASE^-$ to restore the ratio of $VASE^-/VASE^+$ (mean ± SE, n = 15–17).

demonstrated that PQBP1 associates with core spliceosome components and has a protein association network enriched for other splicing factors. Following down-regulation of PQBP1, we observed aberrant splicing profiles for many mRNAs in both human cells and primary mouse cortical neurons. Specifically in neurons, mRNA profiling showed preferential misregulation of AS in factors that are important for neurite outgrowth, which is defective upon PQBP1 depletion. Collectively, we showed that PQBP1 has a role in AS and indicate aberrant AS as a possible mechanism underlying the symptoms of neurological disorders associated with PQBP1.

PQBP1 affects AS

Our work presents several lines of evidence that PQBP1 affects the AS of many pre-mRNAs. Previous work has indicated an association of PQBP1 with mRNA processing factors (Waragai et al. 1999, 2000; Okazawa et al. 2002; Komuro et al. 1999). We discovered that PQBP1 colocalizes with SC35 in nuclear speckles (Fig. 4B; Supplemental Fig. S3A), which are nuclear subdomains enriched for many splicing regulatory factors (Spector and Lamond 2011). Moreover, profiling of PQBP1-associated proteins revealed a set of factors highly enriched in splicing regulators and spliceosome components (Fig. 2E; Supplemental Table S1). Among these binding partners, we also identified WBP11, a previously reported interactor of PQBP1 (Supplemental Table S1; Komuro et al. 1999). These results support PQBP1's connection with the splicing machinery.

PQBP1 affects U2 snRNP component SF3B1 function

We observed that PQBP1 did not associate with functional AS target mRNAs in vivo in CLIP experiments, suggesting that it does not bind directly to the AS targets identified in HeLa cells (Fig. 3A), consistent with its apparent lack of known RNA-binding domains (Fig. 2A). SF3B1, which was previously shown to directly bind substrate mRNAs and affect *Bcl-x* AS (Gozani et al. 1998; Massiello et al. 2006) and was herein identified in PQBP1 affinity purifications (Supplemental Table S1), constituted a prime candidate to work together with PQBP1. Consistent with this hypothesis, depletion of PQBP1 decreased SF3B1's association with *Bcl-x* mRNA (Fig. 3D). In contrast, the binding of SF3B1 to *Mcl1*, a target of SF3B1 (but not of PQBP1) was not affected by the knockdown of PQBP1 (Fig. 3D). These findings suggest that PQBP1 affects the association of SF3B1 with a distinct subset of substrate mRNAs. Furthermore, we found that a subset of PQBP1 targets affected in their splicing was similarly influenced by the knockdown of SF3B1 (Supplemental Fig. S5). SF3B1 is a central component of the U2 snRNP. Recruitment of U2 snRNP to the intronic branchpoint sequence (BPS) is a key step in the assembly of the spliceosomal A complex and thus the selection of splice sites (Wahl et al. 2009). We propose a model whereby through an association with SF3B1, PQBP1 influences the recruitment of U2 snRNP to specific target sites, thus affecting splicing decisions on a subset of mRNAs (Fig. 3E). The precise mechanism by which this occurs remains to be explored.

PQBP1 affects the AS of mRNAs in neurons

Depletion of PQBP1 causes a decrease in outgrowth and branching of dendrites in mouse primary cortical neurons (Fig. 4G,H; Supplemental Fig. S3B,C). This morphological defect was not apoptosis-related (Fig. 4E,F). Qualitatively, dendrites close to the nucleus appeared more affected than distal dendrites by knockdown of PQBP1 (Figs. 4H, 5D). The defect in neurite morphology caused by loss of PQBP1 could be rescued by re-expression of wild-type PQBP1 (Fig. 5C,D), confirming the correlation between neurite deformity and the loss of PQBP1 expression. Transcriptome-wide profiling revealed that PQBP1's AS targets were enriched for factors regulating neuron projection development (Fig. 6C).

To further investigate the effect of PQBP1 on neurons, we focused on the VASE-associated isoforms of NCAM-140 encoded by *Ncam1*, which has been implicated in neurite outgrowth and development. Ncam1 contains five Ig-like domains involved in homophilic binding of Ncam1 either in *cis* or *trans*, which in turn controls cell–cell interactions and the induction of neurite outgrowth (Walsh and Doherty 1997). The AS of VASE determines the inclusion or exclusion of the fourth Ig domain and alters the homophilic binding of NCAM-140 (Lahrtz et al. 1997). The $VASE^-$ isoform of NCAM-140 stimulates—while $VASE^+$ inhibits—neurite outgrowth (Doherty et al. 1992; Liu et al. 1993; Saffell et al. 1994).

We found that the ratio of $VASE^-/VASE^+$ expression corresponded to the level of functional PQBP1 and neurite outgrowth (Figs. 4D, 5C, 7C,E,F). Moreover, restoration of the proper ratio of $VASE^-/VASE^+$ "rescued" the abnormal neurite outgrowth associated with the knockdown of PQBP1 (Fig. 7F,G). These results identified NCAM-140 as a critical target connecting PQBP1-mediated AS to the cellular process of neurite projection and outgrowth. However, the ability of enforced $VASE^-/VASE^+$ expression to rescue the neurite outgrowth defect caused by PQBP1 down-regulation does not exclude the possibility that other PQBP1 targets contribute to this phenotype. For example, PQBP1 AS targets are also enriched in ARF-related proteins (Fig. 6C). The members of the ARF family of small GTPases were previously reported to affect dendritic branching and spine formation (Hernández-Deviez et al. 2002; Moore et al. 2007).

A possible role for PQBP1 in neurological disorders

Clinical genetics identified mutations in PQBP1 as the direct cause for Renpenning syndrome, a representative XLMR disorder characterized by microcephaly, short stature, multiple body malformations, and mental retardation (Kalscheuer et al. 2003; Stevenson et al. 2005). We examined the function of two disease-associated PQBP1 variants—a truncated protein resulting from a ΔAG deletion and a point mutation containing a Y65C mutation in the WW domain. We discovered that the disease-associated PQBP1 mutants had deficiencies in splicing-related functions. Our proteomics profiling indicates that the association between splicing-related factors and wild-type PQBP1 was lost or altered in the Y65C and ΔAG mutants, respectively (Supplemental Table S1). These results indicate that the association of PQBP1 with splicing factors is important for its function. In line with their compromised association with the splicing machinery, neither ΔAG nor Y65C restored the AS of NCAM-140 VASE when PQBP1 was knocked down in mouse primary neurons (Fig. 7C). Furthermore, both mutants were also unable to rescue the defects in neurite outgrowth upon depletion of PQBP1 (Fig. 5E,F). Defects in dendritic development are consistently found in many other mental retardation disorders, including Fragile-X mental retardation and Rett syndrome (Parrish et al. 2007). Our results suggest that disease-associated PQBP1 variants are deficient in maintaining the proper function of PQBP1 in AS. Moreover, aberrant splicing of specific AS targets may contribute to the disease symptoms in PQBP1-related mental retardation disorders.

In summary, we propose that PQBP1, a protein related to neurological diseases, affects AS through association with the core splicing complex. Aberrant AS of specific targets resulting from misregulation of PQBP1 mutants may be one of the causes of PQBP1-related neurological disorders.

Materials and methods

Plasmids, cell culture, and siRNA transfection

Details of plasmid and stable cell line constructions are included in the Supplemental Material. HeLa and 293T cell lines (American Type Culture Collection) were grown in standard conditions. siRNAs were transfected with Hiperfect (Qiagen) at a final concentration of 40 nM. siRNAs were siGENOME siRNAs (Thermo Scientific) for PQBP1 and firefly luciferase. Transfections were optimized to reach ≥85% protein depletion 72 h post-transfection, as determined by Western blotting.

Mouse cortical culture, plasmid transfection, and lentivirus infection

Mouse embryonic cortical neurons were prepared and seeded as in Kim et al. (2010). For direct neuron transfection, MISSION shRNA lentivirus constructs for PQBP1 (Sigma) and a nontargeting shRNA vector (Addgene Plasmid 1864) were transfected into neurons on the day of seeding with Lipofectamine 2000 (Invitrogen). For lentiviral infection of neurons, 293T cells were transfected with shRNA constructs and packaging vectors. Lentivirus-containing supernatants were collected 48 h post-transfection. Neurons were then incubated with the supernatants for 6 h on the day of seeding. Infection was optimized to reach ≥60% protein depletion 72 h post-infection, as determined by Western blotting.

RNA extraction and RT–PCR

Cellular RNA was extracted by RNeasy minikit (Qiagen). RNAs (1.5–4 μg) were reverse-transcribed by SuperScript First-Strand Synthesis system (Invitrogen). *Bcl-x* and NCAM-140 VASE splicing isoforms were analyzed by RT–PCR through primers flanking the alternatively spliced region. A linear range for cDNA input was determined by running a twofold dilution series of 28-cycle PCR reactions. Products were run on 5% Tris-Borate-EDTA gels (Bio-Rad), stained with SYBR gold stain (Invitrogen), and quantified with ImageJ software (National Institutes of

Health). Triplicate samples in the linear range were run to determine isoform ratios. Real-time PCR primers were designed with the software Primer3 (SourceForge) and verified by electrophoresis to generate single, correctly sized products. PCR reactions were prepared with SYBR Green master mix (Applied Biosystems), run on a MasterCycler ep realplex (Eppendorf), and quantified by the $\Delta\Delta C_t$ method. Primer sets are listed in Supplemental Table S5.

Tandem affinity immunopurification

Flag-HA-tagged PQBP1 and the Y65C and ΔAG mutants were purified from HeLa cell extracts by tandem affinity purification and visualized by silver stain as recently described (Adelmant et al. 2012). Specific details are provided in the Supplemental Material.

Sample preparation and mass spectrometry analysis

Sample preparation and liquid chromatography/tandem mass spectrometry analysis were done with slight modifications to a recently described method (Rozenblatt-Rosen et al. 2012). Specific details are provided in the Supplemental Material.

Western blotting

Cellular proteins were extracted with 1× RIPA buffer (1× PBS, 1% Triton X-100, 0.5% sodium deoxycholate, 0.1% SDS, protease inhibitors). Ten micrograms to 30 μg of protein was loaded onto each lane of a NuPAGE 4%–12% Bis-Tris gel. Proteins were transferred onto PVDF membranes and probed with the following primary antibodies diluted in 1× PBST with 5% nonfat milk: α-PQBP1 (Santa Cruz Biotechnology), α-Flag (Sigma), α-HA (Roche), α-actin (Chemicon), and α-SF3B1 (MBL).

CLIP

HeLa cells were cross-linked at 4000 mJ/cm^2 with a UV Stratalinker (Stratgene) and suspended in 1× RIPA buffer supplemented with protease inhibitors, 10 mM DTT, and 0.4 U/μL RNaseOUT (Invitrogen). Lysates were treated briefly with RQ1 RNase-free DNase (Promega) and cleared by centrifugation. Dynal beads (Invitrogen) that were precoupled with 10 μg of antibody or nonrelevant IgG were incubated with the supernatants and rocked for 4 h at 4°C. Beads were then washed three times with the lysis buffer plus 300 mM NaCl and twice with 1× PNK buffer (50 mM Tris-Cl at pH 7.4, 10 mM MgCl2, 0.5% NP-40). RNA–protein complexes were eluted and treated with Protease K (Ambion), and RNAs were extracted by acid phenol/chloroform. Reverse transcription was performed with a 1:1 mix of oligo-dT and random hexamers, and cDNAs were analyzed by real-time PCR.

Immunofluorescence

Neurons were fixed with 4% paraformaldehyde and 2% sucrose/PBS for 8 min. Cells were then washed three times with 1× PBS and treated with 1× GDB (gelatin dilution buffer) buffer (0.3% Triton, 0.1% gelatin) for 15 min. Cells were then incubated with antibodies, stained with DAPI, and imaged on a Nikon Ti inverted fluorescence microscope. For costaining, we swapped the fluorophores on secondary antibodies to confirm there was no bleed-through in the fluorescence channels.

Annexin V staining

Neurons were incubated with FITC-conjugated Annexin V (BioLegend) diluted 1:50 in medium for half an hour at 37°C. Neurons were then directly imaged alive on a Nikon Ti inverted fluorescence microscope.

RNA-seq sample preparation

Libraries were prepared from total RNA following the instructions of the TruSeq RNA sample preparation kit (Illumina). Samples were then run on Illumina HiSeq 2000. Detailed computational methods for AS target profiling are provided in the Supplemental Material. The reported sequencing data have been deposited in NCBI's Gene Expression Omnibus (GEO) (Edgar et al. 2002) and are accessible through GEO series accession number GSE44402.

GO functional analysis

GO term enrichments for proteins associated with PQBP1 in HeLa cells were determined with the DAVID (Database for Annotation, Visualization, and Integrated Discovery) Bioinformatics Database (Dennis et al. 2003; Huang et al. 2009). Functional GO enrichments of PQBP1 AS targets in neurons were determined with GoMiner software (Zeeberg et al. 2003).

Acknowledgments

We thank Hsi-Wen Liao, Athar Malik, David Lipton, and Seth Margolis for help with neuron culture and RNAi; Alan Mardinly for help with Sholl analysis; Jennifer Waters for help with imaging; and Daniel Ducat and Yvonne Chen for thoughtful critiques. We are indebted to Michael Greenberg's laboratory for the generous sharing of mouse embryonic cortical neurons. Microscope imaging was done in the Nikon Imaging Center at Harvard Medical School. Sequencing was performed by the biopolymer facility at Harvard Medical School. This work was supported by grants GM36373 and GM057476 to P.A.S. from the National Institutes of Health.

References

Adelmant G, Calkins AS, Garg BK, Card JD, Askenazi M, Miron A, Sobhian B, Zhang Y, Nakatani Y, Silver PA, et al. 2012. DNA ends alter the molecular composition and localization of Ku multicomponent complexes. *Mol Cell Proteomics* **11:** 411–421.

Akgul C, Moulding DA, Edwards SW. 2004. Alternative splicing of Bcl-2-related genes: Functional consequences and potential therapeutic applications. *Cell Mol Life Sci* **61:** 2189–2199.

Black DL. 2003. Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* **72:** 291–336.

Castle JC, Zhang C, Shah JK, Kulkarni AV, Kalsotra A, Cooper TA, Johnson JM. 2008. Expression of 24,426 human alternative splicing events and predicted *cis* regulation in 48 tissues and cell lines. *Nat Genet* **40:** 1416–1425.

Chakarova CF, Hims MM, Bolz H, Abu-Safieh L, Patel RJ, Papaioannou MG, Inglehearn CF, Keen TJ, Willis C, Moore AT, et al. 2002. Mutations in HPRP3, a third member of pre-mRNA splicing factor genes, implicated in autosomal dominant retinitis pigmentosa. *Hum Mol Genet* **11:** 87–92.

Charizanis K, Lee KY, Batra R, Goodwin M, Zhang C, Yuan Y, Shiue L, Cline M, Scotti MM, Xia G, et al. 2012. Muscleblind-like 2-mediated alternative splicing in the developing brain and dysregulation in myotonic dystrophy. *Neuron* **75:** 437–450.

Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA. 2003. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* **4:** 3.

Doherty P, Moolenaar CE, Ashton SV, Michalides RJ, Walsh FS. 1992. The VASE exon downregulates the neurite growth-promoting activity of NCAM 140. *Nature* **356:** 791–793.

Dredge BK, Polydorides AD, Darnell RB. 2001. The splice of life: Alternative splicing and neurological disease. *Nat Rev Neurosci* **2:** 43–50.

Edgar R, Domrachev M, Lash AE. 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30:** 207–210.

Gehman LT, Stoilov P, Maguire J, Damianov A, Lin CH, Shiue L, Ares M Jr, Mody I, Black DL. 2011. The splicing regulator Rbfox1 (A2BP1) controls neuronal excitation in the mammalian brain. *Nat Genet* **43:** 706–711.

Gozani O, Potashkin J, Reed R. 1998. A potential role for U2AF-SAP 155 interactions in recruiting U2 snRNP to the branch site. *Mol Cell Biol* **18:** 4752–4760.

Grabowski PJ, Black DL. 2001. Alternative RNA splicing in the nervous system. *Prog Neurobiol* **65:** 289–308.

Hernández-Deviez DJ, Casanova JE, Wilson JM. 2002. Regulation of dendritic development by the ARF exchange factor ARNO. *Nat Neurosci* **5:** 623–624.

Huang DW, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4:** 44–57.

Kalscheuer VM, Freude K, Musante L, Jensen LR, Yntema HG, Gécz J, Sefiani A, Hoffmann K, Moser B, Haas S, et al. 2003. Mutations in the polyglutamine binding protein 1 gene cause X-linked mental retardation. *Nat Genet* **35:** 313–315.

Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptewicz M, Barbara-Haley K, Kuersten S, et al. 2010. Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465:** 182–187.

Komuro A, Saeki M, Kato S. 1999. Association of two nuclear proteins, Npw38 and NpwBP, via the interaction between the WW domain and a novel proline-rich motif containing glycine and arginine. *J Biol Chem* **274:** 36513–36519.

Kunde SA, Musante L, Grimme A, Fischer U, Muller E, Wanker EE, Kalscheuer VM. 2011. The X-chromosome-linked intellectual disability protein PQBP1 is a component of neuronal RNA granules and regulates the appearance of stress granules. *Hum Mol Genet* **20:** 4916–4931.

Lahrtz F, Horstkorte R, Cremer H, Schachner M, Montag D. 1997. VASE-encoded peptide modifies NCAM- and L1-mediated neurite outgrowth. *J Neurosci Res* **50:** 62–68.

Licatalosi DD, Darnell RB. 2006. Splicing regulation in neurologic disease. *Neuron* **52:** 93–101.

Liu L, Haines S, Shew R, Akeson RA. 1993. Axon growth is enhanced by NCAM lacking the VASE exon when expressed in either the growth substrate or the growing axon. *J Neurosci Res* **35:** 327–345.

Lunde BM, Moore C, Varani G. 2007. RNA-binding proteins: Modular design for efficient function. *Nat Rev Mol Cell Biol* **8:** 479–490.

Makarova OV, Makarov EM, Urlaub H, Will CL, Gentzel M, Wilm M, Lührmann R. 2004. A subset of human 35S U5 proteins, including Prp19, function prior to catalytic step 1 of splicing. *EMBO J* **23:** 2381–2391.

Massiello A, Roesser JR, Chalfant CE. 2006. SAP155 binds to ceramide-responsive RNA *cis*-element 1 and regulates the alternative 5′ splice site selection of Bcl-x pre-mRNA. *FASEB J* **20:** 1680–1682.

McKie AB, McHale JC, Keen TJ, Tarttelin EE, Goliath R, van Lith-Verhoeven JJ, Greenberg J, Ramesar RS, Hoyng CB, Cremers FP, et al. 2001. Mutations in the pre-mRNA splicing factor gene PRPC8 in autosomal dominant retinitis pigmentosa (RP13). *Hum Mol Genet* **10:** 1555–1562.

Moore CD, Thacker EE, Larimore J, Gaston D, Underwood A, Kearns B, Patterson SI, Jackson T, Chapleau C, Pozzo-Miller L, et al. 2007. The neuronal Arf GAP centaurin α1 modulates dendritic differentiation. *J Cell Sci* **120:** 2683–2693.

Moore MJ, Wang Q, Kennedy CJ, Silver PA. 2010. An alternative splicing network links cell-cycle control to apoptosis. *Cell* **142:** 625–636.

Musante L, Kunde S-A, Sulistio TO, Fischer U, Grimme A, Frints SGM, Schwartz CE, Martínez F, Romano C, Ropers H-H, et al. 2010. Common pathological mutations in PQBP1 induce nonsense-mediated mRNA decay and enhance exclusion of the mutant exon. *Hum Mutat* **31:** 90–98.

Okazawa H, Rich T, Chang A, Lin X, Waragai M, Kajikawa M, Enokido Y, Komuro A, Kato S, Shibata M, et al. 2002. Interaction between mutant ataxin-1 and PQBP-1 affects transcription and cell death. *Neuron* **34:** 701–713.

Parrish JZ, Emoto K, Kim MD, Jan YN. 2007. Mechanisms that regulate establishment, maintenance, and remodeling of dendritic fields. *Annu Rev Neurosci* **30:** 399–423.

Qi Y, Hoshino M, Wada Y-I, Marubuchi S, Yoshimura N, Kanazawa I, Shinomiya K-I, Okazawa H. 2005. PQBP-1 is expressed predominantly in the central nervous system during development. *Eur J Neurosci* **22:** 1277–1286.

Rozenblatt-Rosen O, Deo RC, Padi M, Adelmant G, Calderwood MA, Rolland T, Grace M, Dricot A, Askenazi M, Tavares M, et al. 2012. Interpreting cancer genomes using systematic host network perturbations by tumor virus proteins. *Nature* **487:** 491–495.

Saffell JL, Walsh FS, Doherty P. 1994. Expression of NCAM containing VASE in neurons can account for a developmental loss in their neurite outgrowth response to NCAM in a cellular substratum. *J Cell Biol* **125:** 427–436.

Sholl DA. 1953. Dendritic organization in the neurons of the visual and motor cortices of the cat. *J Anat* **87:** 387–406.

Small SJ, Akeson R. 1990. Expression of the unique NCAM VASE exon is independently regulated in distinct tissues during development. *J Cell Biol* **111:** 2089–2096.

Spector DL, Lamond AI. 2011. Nuclear speckles. *Cold Spring Harb Perspect Biol* **3:** a000646.

Stevenson RE, Bennett CW, Abidi F, Kleefstra T, Porteous M, Simensen RJ, Lubs HA, Hamel BCJ, Schwartz CE. 2005. Renpenning Syndrome comes into focus. *Am J Med Genet* **134:** 415–421.

Tapia VE, Nicolaescu E, McDonald CB, Musi V, Oka T, Inayoshi Y, Satteson AC, Mazack V, Humbert J, Gaffney CJ, et al. 2010. Y65C missense mutation in the WW domain of the Golabi–Ito–Hall syndrome protein PQBP1 affects its binding activity and deregulates pre-mRNA splicing. *J Biol Chem* **285:** 19391–19401.

Ule J, Jensen KB, Ruggiu M, Mele A, Ule A, Darnell RB. 2003. CLIP identifies Nova-regulated RNA networks in the brain. *Science* **302:** 1212–1215.

Ule J, Ule A, Spencer J, Williams A, Hu J-S, Cline M, Wang H, Clark T, Fraser C, Ruggiu M, et al. 2005. Nova regulates brain-specific splicing to shape the synapse. *Nat Genet* **37:** 844–852.

Varani G, Nagai K. 1998. RNA recognition by RNP proteins during RNA processing. *Annu Rev Biophys Biomol Struct* **27:** 407–445.

Vithana EN, Abu-Safieh L, Allen MJ, Carey A, Papaioannou M, Chakarova C, Al-Maghtheh M, Ebenezer ND, Willis C, Moore AT, et al. 2001. A human homolog of yeast pre-mRNA splicing gene, PRP31, underlies autosomal dominant retinitis pigmentosa on chromosome 19q13.4 (RP11). *Mol Cell* **8:** 375–381.

Wahl MC, Will CL, Lührmann R. 2009. The spliceosome: Design principles of a dynamic RNP machine. *Cell* **136:** 701–718.

Walsh FS, Doherty P. 1997. Neural cell adhesion molecules of the immunoglobulin superfamily: Role in axon growth and guidance. *Annu Rev Cell Dev Biol* **13:** 425–456.

Wang G-S, Cooper TA. 2007. Splicing in disease: Disruption of the splicing code and the decoding machinery. *Nat Rev Genet* **8:** 749–761.

Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB. 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature* **456:** 470–476.

Waragai M, Lammers CH, Takeuchi S, Imafuku I, Udagawa Y, Kanazawa I, Kawabata M, Mouradian MM, Okazawa H. 1999. PQBP-1, a novel polyglutamine tractbinding protein, inhibits transcription activation by Brn-2 and affects cell survival. *Hum Mol Genet* **8:** 977–987.

Waragai M, Junn E, Kajikawa M, Takeuchi S, Kanazawa I, Shibata M, Mouradian MM, Okazawa H. 2000. PQBP-1/Npw38, a nuclear protein binding to the polyglutamine tract, interacts with U5-15kD/dim1p via the carboxylterminal domain. *Biochem Biophys Res Commun* **273:** 592–595.

Zeeberg BR, Feng W, Wang G, Fojo AT, Sunshine M, Narasimhan S, Kane DW, Reinhold WC, Lababidi S, Bussey KJ, et al. 2003. GoMiner: A resource for biological interpretation of genomic and proteomic data. *Genome Biol* **4:** R28.

APPENDIX D: SUPPLEMENTAL MATERIALS

SUPPLEMENTAL MATERIALS FOR CHAPTER 2

Supplemental experimental procedures can be accessed from:

http://www.sciencedirect.com/science/article/pii/S0092867410007889

Six supplemental tables and seven supplemental figures as described below can also be accessed from the website above.

Table S2-1: Primary *Bcl-x* whole-genome siRNA screen data.

Table S2-2: *Bcl-x* validation screen data.

Table S2-3: Bcl2-overexpression screening results.

Table S2-4: Results from *Mcl1* counterscreen of *Bcl-x* regulators.

Table S2-5: *Mcl1* and *Bcl-x* kinase/phosphatase comparison.

Table S2-6: Primer sets used in this study.

Figure S2-1: Support Vector Machine (SVM) model for hit-identification in whole-genome siRNA screen.

Figure S2-2: Global analysis of screen assays.

Figure S2-3: Screen hits regulate endogenous *Bcl-x* at physiologically significant levels.

Figure S2-4: Screen hits regulate endogenous *Mcl1* splicing.

Figure S2-5: Pharmacological inhibition of AURKA alters endogenous *Bcl-x* and *Mcl1* splicing.

Figure S2-6: Specific downregulation of ASF/SF2 by AURKA inhibition.

Figure S2-7: Analysis of additional alternative splicing targets.

SUPPLEMENTAL MATERIALS FOR CHAPTER 3

Supplemental Materials and Methods

Plasmid Construction and Cell Lines

Human PQBP1 construct was cloned from the cDNA of HeLa cells. Mouse NCAM-140 VASE – was cloned from the mouse embryonic brain cDNA. ΔAG and Y65C were generated through PCR-mediated mutagenesis. FLAG-HA tandem tags were first cloned into the pBABE vector (BamHI/EcoRI) (Addgene) and PQBP1, ΔAG and Y65C were then cloned into the pBABE-FLAG-HA vector backbone (EcoRI/SalI). For the rescue constructs, puromycin resistance coding sequence was excised from the pLKO.1 lentivirus vector carrying non-targeting shRNA or PQBP1 shRNA (BamHI/KpnI). FLAG-HA tagged human PQBP1, ΔAG, Y65C or mouse NCAM-140 VASE – isoform were then cloned into the vector using isothermal assembly (Gibson et al. 2009). All PCR primers in the cloning steps are listed in Table S3-2.

Stable HeLa cell lines were made by the pBABE-puro retrovirus vectors. 293T cells were transfected with pBABE-puro-FLAG-HA-PQBP1/ΔAG/Y65C constructs together with packaging plasmids. Supernatants were collected after 48 hours and 72 hours. HeLa cells were incubated with virus-containing supernatants for 12 hours. After recovery for 48 hours after the infection, HeLa cells were selected with puromycin and expanded.

Tandem affinity immunopurification

Whole-cell lysates from HeLa cells expressing FLAG-HA dual tagged PQBP1 constructs were prepared in FLAG-IP buffer (Adelmant et al. 2012). Flag-agarose slurry (Sigma)

was incubated with cell lysate for 4 hours at 4°C. After washing and elution with FLAG peptide (Sigma), the eluted products were incubated with HA-agarose slurry (Santa Cruz) overnight at 4°C. The immunopurified products with beads were then washed and eluted with HA peptide (Covance).

Sample preparation for mass-spectrometry analysis

Purified protein complexes were denatured and reduced by incubation at 56°C for 30 min in 10 mM DTT and 0.1% RapiGest (Waters). Protein digestion was carried out overnight at 37°C after adding 2 μg of trypsin and adjusting the pH to 8.0 with 1 M Tris. Tryptic peptide clean-up: RapiGest was removed from solution following the manufacturer's protocol and tryptic peptides were purified by batch-mode reverse-phase C18 chromatography (Poros 10R2, Applied Biosystems) using 40 μL of a 50% bead slurry in RP buffer A (0.1% trifluoroacetic acid), washed with 100 μL of the same buffer and eluted with 50 μL of RP buffer B (40% acetonitrile in 0.1% trifluoroacetic acid). After vacuum concentration, peptides were solubilized in 20 μL of 20 mM sodium phosphate (pH 7.4) and incubated with 20 μL of thiol-activated sepharose 4B (GE-healthcare) to remove excess HA peptide. Potential cysteine-containing peptides recovered in the flow-through were alkylated at room temperature for 20 min in the dark by adding iodoacetamide at a final concentration of 20 mM. Peptides were further purified by strong cation exchange SCX chromatography (Poros 20HS, Applied Biosystems) using 20 μL of a 50% bead slurry in SCX buffer A (25% ACN in 0.1% formic acid), washed with 20 μL of the same buffer and eluted with 20 μL of SCX buffer B (25% ACN, 300 mM KCl in 0.1% formic acid).

LC-MS/MS data acquisition

After vacuum concentration, samples were loaded onto a pre-column (100 µm I.D.; packed with 4 cm POROS 10R2, Applied Biosystems) at a flow rate of 4 µL/min for 15 min using a NanoAcquity Sample Manager (20 µL sample loop) and UPLC pump (Ficarro et al. 2009). After loading, the peptides were gradient eluted (1-30% B in 45 min; buffer A: 0.2 M acetic acid, buffer B: 0.2 M acetic acid in acetonitrile) at a flow rate of ~50 nL/min to an analytical column (30 µm I.D. packed with 12 cm Monitor 5 mm C18 from Column Engineering, Ontario, CA), and introduced into an LTQ-Orbitrap XL mass spectrometer (ThermoFisher Scientific) by electrospray ionization (spray voltage = 2200V). Three rapid LC gradients were included at the end of every analysis to minimize peptide carry-over between successive analyses and to re-equilibrate the columns. The mass spectrometer was programmed to operate in data dependent mode, such that the top 8 most abundant precursors in each MS scan (detected in the Orbitrap mass analyzer, resolution = 60,000) were subjected to MS/MS (CAD, linear ion trap detection, collision energy = 35%, precursor isolation width = 2.8 Da, threshold = 20,000). Dynamic exclusion was enabled with a repeat count of 1 and a repeat duration of 15 sec.

Database searching and identification of the associated protein spectrum for WT PQBP1, ΔAG and Y65C

Orbitrap raw data files were processed within the multiplierz software environment (Parikh et al. 2009). MS spectra were recalibrated using the background ion $Si(CH3)_2O))_6$ at m/z 445.12 +/- 0.03 and converted into a Mascot generic format (.mgf). Spectra were

searched using Mascot version 2.3 against 3 appended databases of i) human protein sequences (downloaded from RefSeq on 07/11/2011) ii) a database of common lab contaminants and iii) a decoy database generated by reversing the sequences from the human database. Search parameters specified a precursor ion mass tolerance of 1 Da, a product ion mass tolerance of 0.6 Da, and methionine oxidation (+16 Da) and cysteine carboxymethylation (+57 Da). The lists of peptide hits from the Mascot searches were filtered to exclude precursors with a mass error greater than 10 ppm. Sequence matches to the decoy database were used to enable a 1% false discovery rate (FDR) filtering of the resulting peptide identifications. After receiving a list of associated proteins for WT PQBP1 and the mutants from database search after LC-MS/MS, we filtered out proteins detected in any one of the replicates from the empty FLAG-HA vector integrated control samples and define the rest as the proteins that associated with WT PQBP1, ΔAG and Y65C.

Antibodies

Antibodies used in this study are listed as follows: α-PQBP1 (Santa Cruz); α-FLAG (Sigma); α-HA (Roche); α-actin (Chemicon); α-SF3B1 (MBL); α-MAP2 (Sigma); α-SC35 (Abcam).

Supplemental Figures

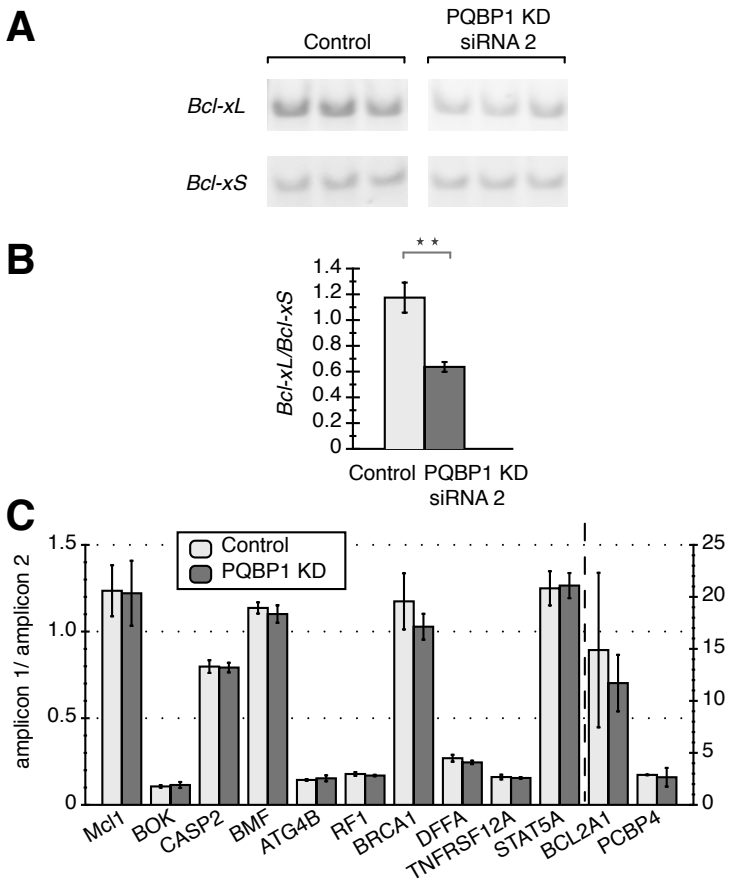

Figure S3-1

(A) Gel analysis of RT-PCR of *Bcl-x* splicing isoforms for samples treated with a control siRNA targeting firefly luciferase and a second siRNA targeting PQBP1 (PQBP1 KD siRNA 2 in Figure 3-1B) is shown.

(B) Quantification of gel analysis as in (A). Mean of three independent measurements +/- SD are shown. Changes were tested by 1-way ANOVA. **: statistically significant with p-value<0.01.

(C) Quantification of real-time PCR data on 12 mRNAs that showed no significant AS changes between control and PQBP1 KD samples; *APAF1*, *TNFSF13*, *GSTCD* and *PKM2* were also checked but showed either no or unstable readout, so are not included here. Y-axis depicts the ratio between amplicons for constitutive and alternative exonic regions; *BCL2A1* and *PCBP4* follow the y-axis on the right side while the rest follow the y-axis on the left side. Light columns show the ratio of two amplicons from the control sample and dark columns from the PQBP1 KD sample. Mean of three independent measurements +/- SD are shown. Difference of the ratio between control and PQBP1 KD samples was analyzed by 1-way ANOVA test.

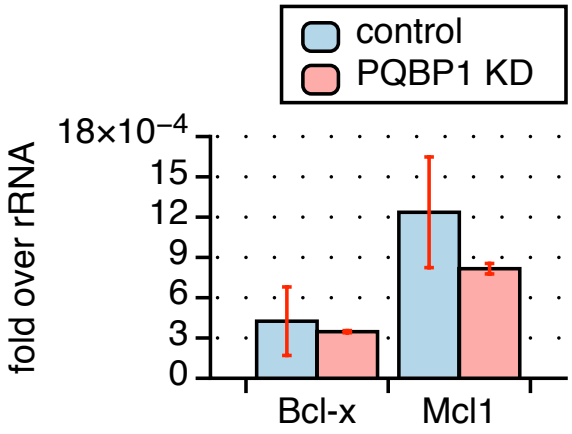

Figure S3-2.

Quantification of steady state mRNA level of *Bcl-x* and *Mcl1* in control and PQBP1 KD samples by RT-PCR. Data values normalized to rRNA expression and mean of three independent measurements +/- SD are shown.

Figure S3-3

(A) IF staining of two additional rat embryonic cortical neurons for PQBP1 and SC35. Neurons were fixed and stained with antibodies listed on top of each panel.

(B) Western blot showing depletion of PQBP1 in mouse embryonic cortical neurons infected by virus from the PQBP1-targeting shRNA in Figure 3-4C (PQBP1 KD shRNA 1), a second PQBP1-targeting shRNA (PQBP1 KD shRNA 2) and a non-targeting control shRNA.

(C) IF staining of the morphology of mouse primary embryonic cortical neurons 5 days after infection with virus from a non-targeting control shRNA and a second PQBP1-targeting shRNA (PQBP1 KD shRNA 2 as in (A)). Green: MAP2 staining; blue: DAPI staining.

(D) Western blot showing that the expression of shRNA-resistant human PQBP1 is not depleted by co-expression of the shRNA targeting mouse endogenous PQBP1. Lane 1 and 2 are samples co-transfected with a GFP-tagged mouse PQBP1 (mPQBP1-GFP) and a non-targeting control shRNA or a PQBP1-targeting shRNA. Lane 3 and 4 are samples transfected with a GFP-tagged, shRNA-resistant human PQBP1 (resis_hPQBP1-GFP) and a non-targeting control shRNA or a PQBP1-targeting shRNA. Antibodies used are listed on the right. GAPDH serves as the loading control.

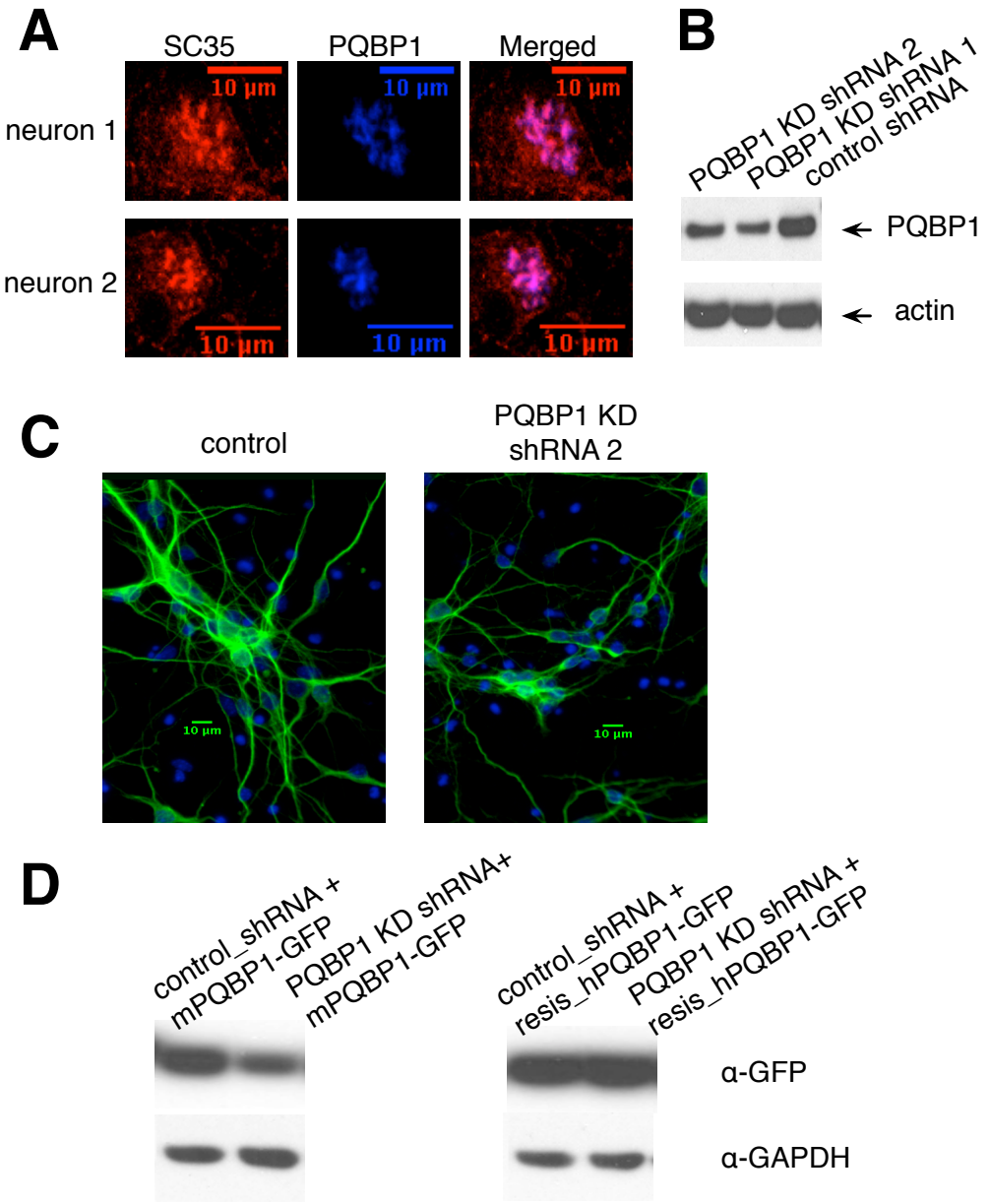

Figure S3-3 (continued).

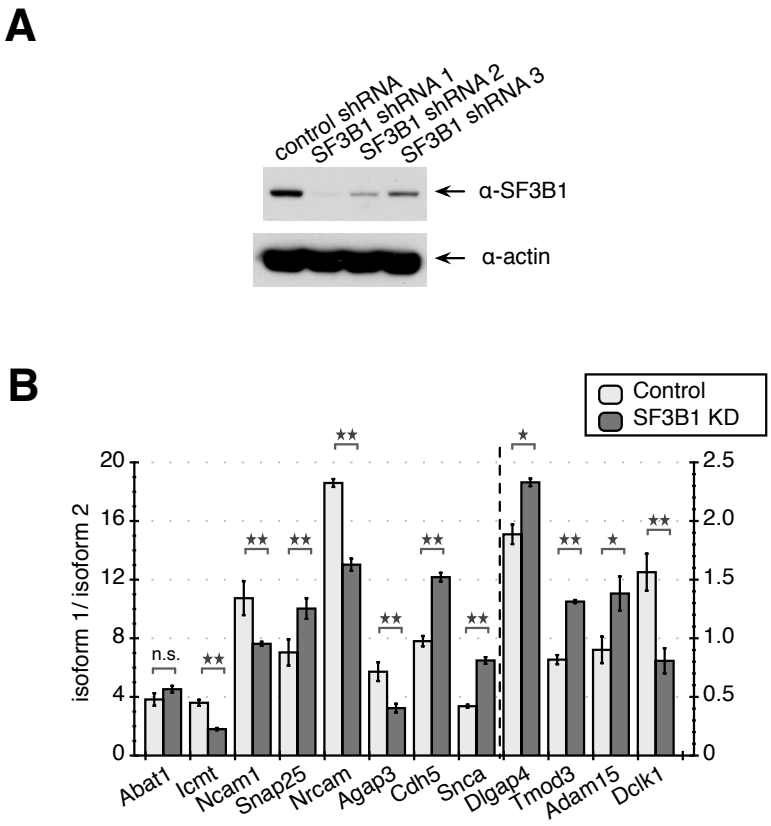

Figure S3-4

(A) Western blot of mouse embryonic cortical neurons infected by virus from a non-targeting control shRNA and three different SF3B1-targeting shRNAs. Antibodies used are listed on the right. Actin serves as the loading control.

(B) RT-PCR quantification of the AS change of 10 verified PQBP1-modulated AS events upon SF3B1 depletion (by shRNA 2 in Figure S3-4A), as described in Figure 3-6B.

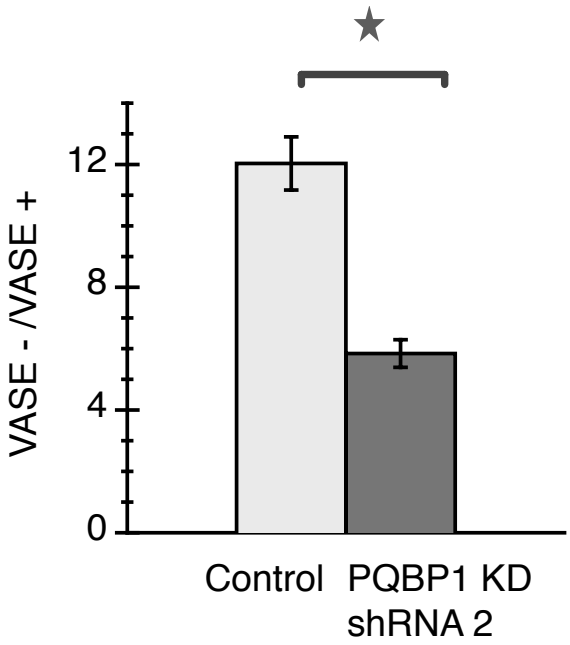

Figure S3-5

Real-time PCR quantification for VASE – and VASE + upon depletion of PQBP1 by a second shRNA is shown. The ratio of the two isoform expression was compared between neurons infected with virus from a non-targeting control shRNA and a second PQBP1-targeting shRNA. Mean of three independent measurements +/- SD are shown. Changes were tested by 1-way ANOVA. ∗: statistically significant with p-value<0.03.

Supplemental Tables

All the supplemental tables described below can be accessed through the link:

http://www.ncbi.nlm.nih.gov/pubmed/23512658

Table S3-1. GO term enrichment for PQBP1 AS targets in mouse embryonic cortical neurons.

The following data are enclosed: the GO category name; total number of genes in the corresponding GO category associated with the background gene set; number of genes within PQBP1 AS targets in the corresponding GO category; enrichment—the proportion of genes from PQBP1 AS targets in the category relative to the expected proportion; the one-sided Fisher exact p value corrected for multiple comparisons (fdr, false discovery rate). This table is generated by GoMiner™ online software (Zeeberg et al., 2003).

Table S3-2. Primer sets used in this study.

PCR primer sequences are listed here. Abbreviations are as follows: F=forward (5') primer; R=reverse (3') primer; A=amplicon for alternative regions; C=amplicon for constitutive regions.

Supplemental references

Adelmant G, Calkins AS, Garg BK, Card JD, Askenazi M, Miron A, Sobhian B, Zhang Y, Nakatani Y, Silver PA, et al. 2012. DNA ends alter the molecular composition and localization of ku multicomponent complexes. *Mol Cell Proteomics* **11:** 411-421.

Ficarro SB, Zhang Y, Lu Y, Moghimi AR, Askenazi M, Hyatt E, Smith ED, Boyer L, Schlaeger TM, Luckey CJ, et al. 2009. Improved electrospray ionization efficiency compensates for diminished chromatographic resolution and enables proteomics analysis of tyrosine signaling in embryonic stem cells. *Anal Chem* **81:** 3440-3447.

Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, Smith HO. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Meth* **6:** 343-345.

Parikh JR, Askenazi M, Ficarro SB, Cashorali T, Webber JT, Blank NC, Zhang Y, Marto JA. 2009. multiplierz: an extensible API based desktop environment for proteomics data analysis. *BMC Bioinformatics* **10:** 364.

Zeeberg BR, Feng W, Wang G, Fojo AT, Sunshine M, Narasimhan S, Kane DW, Reinhold WC, Lababidi S, Bussey KJ, et al. 2003. GoMiner: a resource for biological interpretation of genomic and proteomic data. *Genome Biol*. **4:** R28.

SUPPLEMENTAL MATERIALS FOR CHAPTER 4

All the supplemental tables described below can be accessed through the link:

http://www.ncbi.nlm.nih.gov/pubmed/23512658

Table S4-1. AS targets of PQBP1 identified from RNA-seq.

The following data for AS events that experienced significant changes upon PQBP1 KD in mouse embryonic cortical neurons are enclosed: Gene name that the AS event belongs to; the AS event ID; sub-AS-junction IDs in this AS event; the corresponding genome coordinates for the sub-AS-junctions; p-values of differential usage of the corresponding sub-AS-junction between control and PQBP1 KD samples (NA p-value: no statistical test applied because of low read count. See text); fold changes of usage (on log2 scale) for the sub-AS-junction between control and PQBP1 KD samples.

Table S4-2. AS events that are not affected by PQBP1 KD from RNA-seq.

AS events that are not affected by PQBP1 KD in mouse embryonic cortical neurons are enclosed, with similar format as in Table S4-1. AS events that could not be mapped to any known genes based on current annotation of the mouse genome are not included here.

SUPPLEMENTAL MATERIALS FOR CHAPTER 5

Supplemental Table S5-1: Parameters used for modeling the splicing-based device.

Parameters	Value
k_1	0.02
k_2	6.66×10^{-7}
l_2	10^{-2}
k_3	1.33×10^{-6}
l_3	10^{-3}
k_4	1
k_5	0.015
d_1	10^{-4}
d_2	5×10^{-4}
d_3	5×10^{-3}